DEA-based performance measurement under centralized management

Von der

Carl-Friedrich-Gauß Fakultät

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

genehmigte Dissertation

von

Sören Guntram Harms

geboren am 16. April 1990

in Braunschweig

Eingereicht am:	17. April 2019
Disputation am:	15. Juli 2019
1. Referent:	Prof. Dr. Heinz Ahn
2. Referent:	Prof. Dr. Thomas Spengler

2019

Acknowledgements

First, I want to thank my supervisor and tutor Prof. Dr. Heinz Ahn not only for his continuous and friendly support during these four years of research, but also for all the freedom I enjoyed during my time at his institute. This gave me the opportunity to work (successfully and unsuccessfully) on different research fields and make my PhD time as educational as it can possibly be. I also want to thank Prof. Dr. Thomas Spengler for being the second reviewer of this dissertation. I am also very grateful for the friendly collaboration with Dr. Mohsen Afsharian. This work would not have been possible without his contributions and good advice. His enthusiasm and extensive knowledge have been enormously helpful and very inspiring for me. I also want to thank all my colleagues from the Institute of Management Control & Business Accounting (TU Braunschweig) for all the fun moments we had at Christmas parties, excursions, conferences and on every busy day of the week. Furthermore, I want to mention the continuous support from my mother and father through all my years of study (before and during my PhD time).

I always must remember what Prof. Ahn said at one of our seminars: “One important thing that every student requires to finish a PhD thesis is perseverance.” After four years of work, I can definitely say: He is correct! A dissertation comes along with seemingly endless literature work, various result discussions and numerous revision loops. However, finishing a PhD thesis also requires – especially as a young father – a helpful and eager partner. To that end, I want to express my infinite love and deep gratitude to my wife *Lisa*. Not only for giving me the necessary freedom to finish this thesis but also for making my PhD time more meaningful. This work is dedicated to her.

Zusammenfassung

Diese Arbeit entstand vor dem Hintergrund einer praktischen Herausforderung bei dem Unternehmen KONE – einem Marktführer in der weltweiten Aufzugs- und Fahrtreppen-Industrie. In Deutschland hat KONE rund 70 lokale Serviceniederlassungen gegründet, welche hauptverantwortlich für die Wartung und Reparatur von Aufzügen und Fahrtreppen in ihren umliegenden Regionen sind. Die verschiedenen Serviceniederlassungen werden gemäß ihrer geographischen Lage in vier unterschiedliche Managementbereiche eingeteilt. Jeder Managementbereich wird von einer zentralen Entscheidungsinstanz verantwortet. Diese Instanz verfügt über die Befugnis eigenständige Strategien zu implementieren, Ressourcenumverteilungen vorzunehmen sowie niederlassungsübergreifende Kooperationen umzusetzen. Diese Flexibilität ermöglicht es den zentralen Entscheidungsinstanzen, das operative Geschäft derart auszugestalten, dass die Ansprüche der lokalen Kundengruppen bestmöglich erfüllt werden.

Durch den starken Preiswettbewerb in der Aufzugs- und Fahrtreppen-Industrie ist KONE seit mehreren Jahren dazu gezwungen, die Performance seines oben beschriebenen Servicesegments kontinuierlich zu verbessern. In diesem Zusammenhang hat KONE u. a. diverse monetäre als auch nicht-monetäre Performancemaße zur Steuerung der Serviceniederlassungen und Managementbereiche etabliert. Wie jeder traditionelle Performance Measurement Ansatz geht auch das bei KONE eingesetzte Kennzahlensystem mit einigen Herausforderungen einher. Beispielsweise ist für eine Aggregation der unterschiedlichen Kennzahlen zu einem einzelnen (übergeordneten) Performancemaß eine vorherige Festlegung von Indikatoren-Gewichtungen notwendig. Dies wiederum erfordert subjektive Wertungen über die jeweilige Bedeutung der einzelnen Kennzahlen. Daraus resultieren konsequenterweise Diskussionen über alternative Gewichtungen – insbesondere mit Vertretern der Serviceniederlassungen und den zentralen Entscheidungsinstanzen. Darüber hinaus

lassen sich in derart traditionellen Performance Measurement Ansätzen nur schwer etwaige Verbesserungspotentiale berücksichtigen, die aus Ressourcenumverteilungen oder aus niederlassungsübergreifenden Kooperationen resultieren. Derartige Einflussgrößen sind allerdings für die Gewinnung von aussagekräftigen Analyseergebnissen als auch für die Bestimmung von realistischen und motivierenden Zielvorgaben substantiell.

Eine betriebswirtschaftliche Methode, welche die oben genannten Limitationen traditioneller Performance Measurement Ansätze nicht aufweist, ist die von Charnes et al. (1978) entwickelte Data Envelopment Analysis (DEA). Diese Methode nutzt eine modell-endogene Gewichtung, um eine Aggregation von mehreren Indikatoren zu einem einzelnen Performancemaß zu ermöglichen. Dadurch entfällt die Notwendigkeit einer vorherigen Festlegung von Kennzahlen-Gewichtungen. Darüber hinaus zeigen neuere Veröffentlichungen, wie sich zentralisierte Managementstrukturen in einer DEA-basierten Performanceanalyse abbilden lassen.

Aufgrund dieser Vorteile wird ein umfassender Literaturüberblick erarbeitet, wie unterschiedliche Zentralisationsgrade auf dem Forschungsgebiet der DEA modelliert werden. Mithilfe der durchgeführten Literaturrecherche wurden insgesamt 135 unterschiedliche Ansätze ermittelt, die (implizit oder explizit) entweder ein vollkommen zentralisiertes oder teilweise zentralisiertes Managementmodell (sog. hybrides Management) unterstellen. Entsprechend der von den jeweiligen Autoren verfolgten Forschungsziele wurden die verschiedenen Veröffentlichungen in acht Themenfeldern klassifiziert. Für jedes Forschungsfeld wurden die einflussreichsten DEA-Ansätze (gemessen an den erhaltenen Zitationen) mathematisch beschrieben und näher erläutert. Etwaige Zusammenhänge zwischen Veröffentlichungen und Themenfeldern wurden mithilfe unterschiedlicher Zitationsanalyse-Techniken untersucht.

Eine abschließende Diskussion der verschiedenen DEA-Ansätze zeigte, dass keine der bisher publizierten Methoden für den speziellen Fall von KONE anwendbar ist. Vor diesem Hintergrund verfolgt diese Arbeit zwei fundamentale Ziele: Einerseits soll ein DEA-basierter Performance-Measurement-Ansatz erarbeitet werden, der zur Messung von Effizienzveränderungen von einzelnen Serviceniederlassungen über die Zeit geeignet ist. Andererseits soll eine weitere DEA-basierte Methode entwickelt werden, welche bei Performancevergleichen zwischen unterschiedlichen Managementgruppen anwendbar ist.

Die aufgrund dieser Überlegungen entstandenen DEA-Ansätze basieren auf der Kombination des sog. Metafrontier-Konzepts mit dem Malmquist-Produktivitätsindex. Der erste hier entwickelte Ansatz erlaubt es, Performanceveränderungen von einzelnen Produktiveinheiten über mehrere Zeitperioden zu messen und gleichzeitig potentielle Ursachen für Performanceveränderungen zu identifizieren. Im Gegensatz zu konventionellen Metafrontier-basierten Malmquist-Indizes berücksichtigt der vorgeschlagene Ansatz die individuellen Eigenschaften der lokalen Produktionstechnologien. Dadurch lassen sich zusätzliche Informationen über die analysierten Managementgruppen gewinnen, die als Ausgangspunkt für weitere Performanceanalysen dienen können.

Der zweite in dieser Arbeit vorgeschlagene DEA-Ansatz nutzt den Malmquist-Produktivitätsindex für den Performancevergleich von Managementgruppen. Der Index berücksichtigt dabei explizit, dass eine zentrale Entscheidungsinstanz existiert, welche u. a. Ressourcenumverteilungen durchführen kann, um die Gesamtperformance des jeweiligen Managementbereichs zu verbessern. Ferner erfüllt der vorgeschlagene Index die Zirkularitätsbedingung und kann dadurch u. a. für die Ableitung von Performance-Rankings genutzt werden.

Abstract

This thesis is motivated by the special case of KONE Corporation – a service and engineering company that is widely recognized as one of the global leaders in the elevator and escalator industry. Throughout Germany, KONE established around 70 local maintenance units, which are responsible for the maintenance and repair of elevators and escalators within their respective geographical area. In order to oversee its maintenance units in an efficient way, KONE has segregated them into four different management groups. Each group is administered by a central decision maker who has the authority to apply customized strategies, undertake resource reallocations or promote collaborations between different maintenance units. In this way, the regional management teams can run the business according to the respective demands of their local costumers.

The fierce price competition in the elevator and escalator industry forced KONE to continuously improve the performance of its service business. In order to monitor the performance of its maintenance units and management groups, KONE applies a comprehensive set of different financial and non-financial indicators. However, like any other traditional performance measurement approach, KONE's framework faces different limitations. Among other things, the applied method requires the previous determination of a set of fixed weights to aggregate the different indicators to an overall performance score. These weights are typically based on value judgements and, hence, cause numerous discussions with the respective representatives of KONE's maintenance units or management groups about alternative weighting schemes. Furthermore, the current performance measurement approach is not able to incorporate additional improvement potentials that can be received from the ability of the regional management to reallocate resources or promote collaborations. However, this aspect is tremendously important to obtain meaningful performance scores and motivational as well as realistic target values.

A framework which is able to overcome the aforementioned limitations of KONE's performance measurement approach has been proposed by Charnes et al. (1978) and is called data envelopment analysis (DEA). This method uses a model endogenous weighting to receive a single overall performance score and, therefore, does not require a previous determination of indicator weights. Furthermore, a variety of recent publications has proven that DEA is also adequate for modeling the improvement potential that can be gained from a centralized management structure.

Against this background, the thesis provides a thorough overview of how different degrees of centralization are modeled in the current DEA literature. The systematic literature review identified 135 different approaches that (implicitly or explicitly) assume a centralized or partially centralized management structure (so-called hybrid management). According to the respective objectives of each publication, the approaches were categorized into eight distinct research streams. Furthermore, the most influential DEA approaches were mathematically described and discussed in greater detail. The interdependencies between the different publications and research streams were examined using different techniques of citation-based analysis.

A concluding discussion of the respective DEA approaches showed that none of the frameworks introduced so far could be directly applied to the practical case of KONE Corporation. In response to this research gap, this thesis has two fundamental objectives: The first objective is to propose a DEA-based performance measurement approach for measuring performance changes of the individual maintenance units over time. The second objective is to develop another DEA-based approach for comparing the performance of the different regional management groups.

Both DEA approaches thus developed are based on the combination of the so-called metafrontier concept and the Malmquist productivity index. The first approach evaluates productivity changes of operating entities over time and, hence, may indicate potential sources for performance changes. Thereby, the proposed approach preserves the individual characteristics of each local group technology – a unique feature which is not shared by conventional metafrontier Malmquist index approaches. In other words, the new method can provide valuable economic information, e.g., about superior management styles. Such information can serve KONE as a possible starting point for a more detailed performance analysis of outperforming maintenance units.

The second DEA approach proposed here introduces a new index for comparing the performance of management groups. This index accounts for the existence of a central decision maker who can, e.g., undertake resource reallocations to improve the overall performance of its managed group. Furthermore, the resulting index satisfies the circularity property, which allows consistent performance rankings to be derived.

The applicability and usefulness of both proposed approaches is empirically shown with real-world data from KONE Corporation.

Table of contents

List of tables

List of figures

List of abbreviations

AFI	Adjusted frontier index
BCC	Banker, Charnes, Cooper (DEA model with VRS)
BPC	Best practice change
CAN	Canada
CCR	Charnes, Cooper, Rhodes (DEA model with CRS)
CD	Camanho, Dyson (Extension of the Malmquist index)
CHN	China
CPI	Centralized performance index
CRS	Constant returns to scale
CSW	Common set of weights
DA	Discriminant analysis
DEA	Data envelopment analysis
DMU	Decision making unit
EC	Efficiency change
EI	Efficiency index
ERP	Enterprise resource planning
FI	Frontier index
FTE	Number of full-time equivalent employees
IRN	Iran
MC	Material costs
MI	Malmquist index
NOC	Number of callouts
PI	Performance index according to Camanho and Dyson (2006)
PPS	Production possibility set
RD	Ray, Desli (Decomposition of the Malmquist index)
REV	Revenue
SFA	Stochastic frontier analysis

TC	Technology change
TGI	Technology gap index
THT	Total handling tasks
VRS	Variable returns to scale
WRT	Weighted response time

1 Introduction

1.1 Background

Until the 1980s, the majority of applied performance measurement approaches evaluated companies in accordance with the main shareholder interest of profit maximization. Accordingly, the respective approaches were predominantly based on financial indicators such as return on investment, returns on sales or sales per employee. However, in response to the increasingly globalized markets and the associated appearance of new competitors from other geographical regions, companies gradually began to change their priorities in the late 1980s (see Ghalayini and Noble 1996). Beside the satisfaction of shareholder interests, customers' and employees' requirements also received increasing attention from the companies' management (see Kaplan and Norton 1996a). The rising significance of these stakeholder groups led to adapted business strategies to withstand growing market challenges such as intensified price pressure and, hence, to ensure long-term financial success (see Ghalayini and Noble 1996).[1]

Due to these fundamental changes at the end of the 1980s, the exclusive application of financial performance measures was increasingly considered inappropriate to supplement the management's decision making processes. Financial indicators are lagging metrics and, therefore, measure the results of decisions in the past. They can only partially reflect

---

1 The interested reader is referred to Barney and Wright (1998) for the importance of employee satisfaction for gaining competitive advantage. For a discussion on customer satisfaction and competitive advantage, see Woodruff (1997).

the interests of customers or employees and, therefore, do not straightforwardly allow predictions about a company's future success (see Ghalayini and Noble 1996).[2]

In response to these limitations of financial indicators, a variety of alternative performance measurement approaches have been developed since the 1980s.[3] The new approaches account for profit maximization interests, but also consistently reflect customers' and employees' demands through the application of numerous non-financial measures (see Howell and Soucy 1987, Vollmann 1988, Dent 1990, Kaplan and Norton 1996b). Widely known representatives of this new category of performance measurement approaches are the SMART system, which was originally developed by Wang Laboratories,[4] the Performance Measurement Framework proposed by Keegan et al. (1989), the Performance Measurement Questionnaire of Dixon (1990) and the Balanced Scorecard of Kaplan and Norton (1996a).

Besides the application of a comprehensive indicator set, there is also widespread scientific agreement that performance measurement approaches need to be tailored according to the respective organizational requirements (see e.g., Globerson 1985, Maskell 1991, Neely et al. 1995).[5] This includes important strategic considerations, also concerning the appropriate organizational structures.[6] To gain a better understanding of this idea, consider the following example:

Consider a company with a variety of organizational units. When evaluating the performance of such an individual entity and comparing it with a benchmarking set, one needs

---

2 A thorough discussion of the major problems of financial indicators has been published by Ittner and Larcker (1998).

3 A comprehensive overview of different performance measurement approaches can be found in Neely et al. (1995).

4 Cross and Lynch (1988) provide a detailed description of this performance measurement approach.

5 Important considerations for designing a performance measurement approach have been mentioned by, e.g., Globerson (1985), Crawford and Cox (1990), Blenkinsop and Davis (1991), Maskell (1991) as well as Wisner and Fawcett (1991).

6 While strategic considerations are repeatedly described in the current literature (see e.g., Atkinson et al. 1997, Kaplan 2001, Ittner et al. 2003), there is only little research regarding the importance of organizational structures for the design of performance measurement approaches. However, some implications have been mentioned by Neely et al. (1995, p. 102), Mar-Molinero et al. (2014) and Afsharian et al. (2019c).

to account for the given organizational structures in order to identify any restrictions regarding a unit's decision making opportunities. In line with this, classifying a certain entity as efficient just because of its concentration on, e.g., certain products, services or customers may be inappropriate when the top management does not favor the specializations mentioned. In such cases, the entities are required to follow a more balanced strategy, which needs to be considered by the applied measurement approach.[7] Also, the individual evaluation of entities would be inappropriate when only a small degree of decision making authority is given to them. If the top management retains, e.g., the right to take significant strategic decisions such as resource reallocations or closing certain service lines or branches, the evaluation should address the entities as a whole to reveal the respective potential for improvement.

From the aforementioned discussions, it can be concluded that performance scores are only meaningful when the applied approach not only incorporates financial and non-financial indicators simultaneously, but also accounts for the respective particularities of the organization. These two requirements to a modern performance measurement approach are essential for effective decision making and the comprehensive detection of potentials for improvement. Against this background, the development of a performance measurement approach, which accounts for both aspects, is the major subject of this thesis.

1.2 Motivation and objectives

This thesis is motivated by the case of KONE Corporation, which is widely recognized as one of the global leaders in the elevator and escalator industry (see KONE 2018b). KONE generates the major share of its yearly revenue with its new equipment business (i.e., the sale of elevators and escalators). However, this business stream is highly dependent on developments in the construction industry and, therefore, follows cyclical

[7] The problem of classifying a certain unit as efficient because of its concentration on certain variables has been addressed by e.g., Dyson and Thanassoulis (1988) as well as Roll et al. (1991). Solutions to this problem have been proposed by Ahn et al. (2012), Dyckhoff et al. (2013) and Dyckhoff and Gutgesell (2015).

fluctuations. In addition, the new equipment business only yields relatively low profit margins and showed a declining sales volume within the last three years (see KONE 2018a). By contrast, KONE's business with maintenance and repair services is mainly independent from other industries and showed increasing revenues in recent years (see KONE 2018a). Furthermore, the service segment usually yields higher profit margins (see VDMA e.V. 2018). Hence, KONE and other members of the elevator and escalator industry tend to accept lower (and even unprofitable) prices for new lift systems when the transactions are combined with profitable long-term service contracts (see VDMA e.V. 2018).

However, price pressure also increased in the service segment over the past couple of years due to higher market power of the elevator and escalator operators (such as national supermarket chains or government authorities). Against this background, KONE is forced to continuously improve the performance of its so-called maintenance units. These operating entities are responsible for the maintenance and repair of elevators and escalators within their defined geographical regions. To better oversee the maintenance units, the German headquarters of KONE has partitioned them into four distinct managerial groups whereby each group is controlled by a central decision maker who has the ability to undertake resource allocations and promote group-wide collaborations.

In the process of performance improvement, the application of an appropriate performance measurement framework is essential. For the special case of maintenance organizations, numerous approaches have been proposed (see e.g., Groote 1995, Kutucuoglu et al. 2001, Parida and Kumar 2006, Parida and Chattopadhyay 2007 and Muchiri et al. 2011). Most researchers agree that measuring performance of maintenance organizations is rather complex and should include numerous financial and non-financial indicators, which is consistent with the discussions in Section 1.1.

Popular approaches such as the cost-benefit analysis use a weighting scheme to aggregate different financial and non-financial indicators to an overall performance score (see Thormählen 1977). However, it is usually criticized that such weightings schemes are highly subjective and, hence, may lead to flawed decision making. Such aggregations may be necessary when a company seeks to rank the different maintenance units and identify performance outliers for a detailed process analysis. Furthermore, these approaches usually cannot account for the complex interdependences between the different variables included in the operational processes of the maintenance units (see Thanassoulis 2001, p. 6). However, the consideration of such interdependencies may be important – especially when a company seeks to set binding and realistic performance targets. Hence, it is questionable that the traditional approaches can sufficiently measure performance and identify the full improvement potentials of KONE's maintenance units.

A promising alternative, which has been introduced by Charnes et al. (1978) based on the seminal work of Farrell (1957), is data envelopment analysis (DEA). The major advantage of this method is that, unlike fixed weight approaches, it does not require numerous a priori assumptions and calculations (see Cooper et al. 2006). Instead, this method uses a so-called "model endogenous weighting" which means that the different indicator weights are an outcome of the computational processes. In addition, missing weight restrictions of basic DEA models ensure that the entity being evaluated cannot improve its received performance score by choosing a different set of weights (see Cooper et al. 2006). This in turn means that discussions with the respective management about an alternative weighting scheme become obsolete (see Ahn 2014). Another attractive feature of DEA is its ability to account for the impact of several contextual factors such as returns to scale. It is therefore not surprising that DEA has been used to measure the performance of operating entities in a variety of different situations. Successful applications have been reported, e.g., for the evaluation of bank branches (see Sherman and Gold 1985), farms (see Fraser and Cordina 1999), hospitals (see Jacobs 2001), countries (see Despotis 2005), electric power plants (see Vaninsky 2006) and numerous other cases.[8]

Also the literature on maintenance performance management considers DEA as "an appropriate method for the quantitative comparison of maintenance organizations" (see Garg and Deshmukh 2006, p. 223). In line with this, different publications have suggested the usefulness of DEA in maintenance-related contexts. For example, Bowlin (1987), Charnes et al. (1984), Roll et al. (1989), Clarke (1992) and Sun (2004) have applied DEA to operating units which are responsible for maintenance services in the military sector.

[8] A thorough overview of different application areas of DEA has been given by Emrouznejad et al. (2008).

Other researchers such as Cook et al. (1991), Cook et al. (1994) as well as Hjalmarsson and Odeck (1996) have applied DEA to evaluate the efficiency of different entities in the road maintenance and construction sector.

Against this background, it is the fundamental objective of this thesis to develop a DEA-based performance measurement approach to appropriately measuring the performance for cases such as KONE's local maintenance units. In contrast to the above-mentioned publications, this thesis will not simply apply basic DEA models. Instead, the proposed performance measurement framework seeks to incorporate the particular structure of the focal organization into its DEA models. This is justified with the tremendous impact of organizational structures on behavior patterns, decision opportunities and corresponding potentials for performance improvement.[9]

Since there are numerous performance measurement approaches and potential organizational structures to be modeled, it is necessary to set priorities and identify those topics, which are of particular relevance for the KONE case. In consideration of the given organizational background and discussions with four representatives of KONE, this thesis will focus on two research streams, which are briefly described below:

1. In order to compare the individual maintenance units of KONE over several time periods, a DEA-based performance index is proposed. In contrast to existing approaches, the suggested methodology explicitly dictates that the maintenance units are organized in distinct management groups and, therefore, face differing production opportunities. Consequently, the approach is based on more accurate assumptions and, hence, will typically lead to more accurate performance scores.
2. This thesis also introduces a new index to compare the performance of groups of operating entities using DEA. The proposed method accounts for improvement potentials that can be received from centrally coordinated resource reallocations or group-wide collaborations. This makes the new approach also applicable to the practical case of KONE´s management groups where a central decision maker oversees a set of subordinated maintenance units.

---

9 The influence of the organization structure on the performance of a company has been described by Armour and Teece (1978) and Olson et al. (2005).

1.3 Outline

In order to meet the previously mentioned objectives, the following chapters unfold as follows:

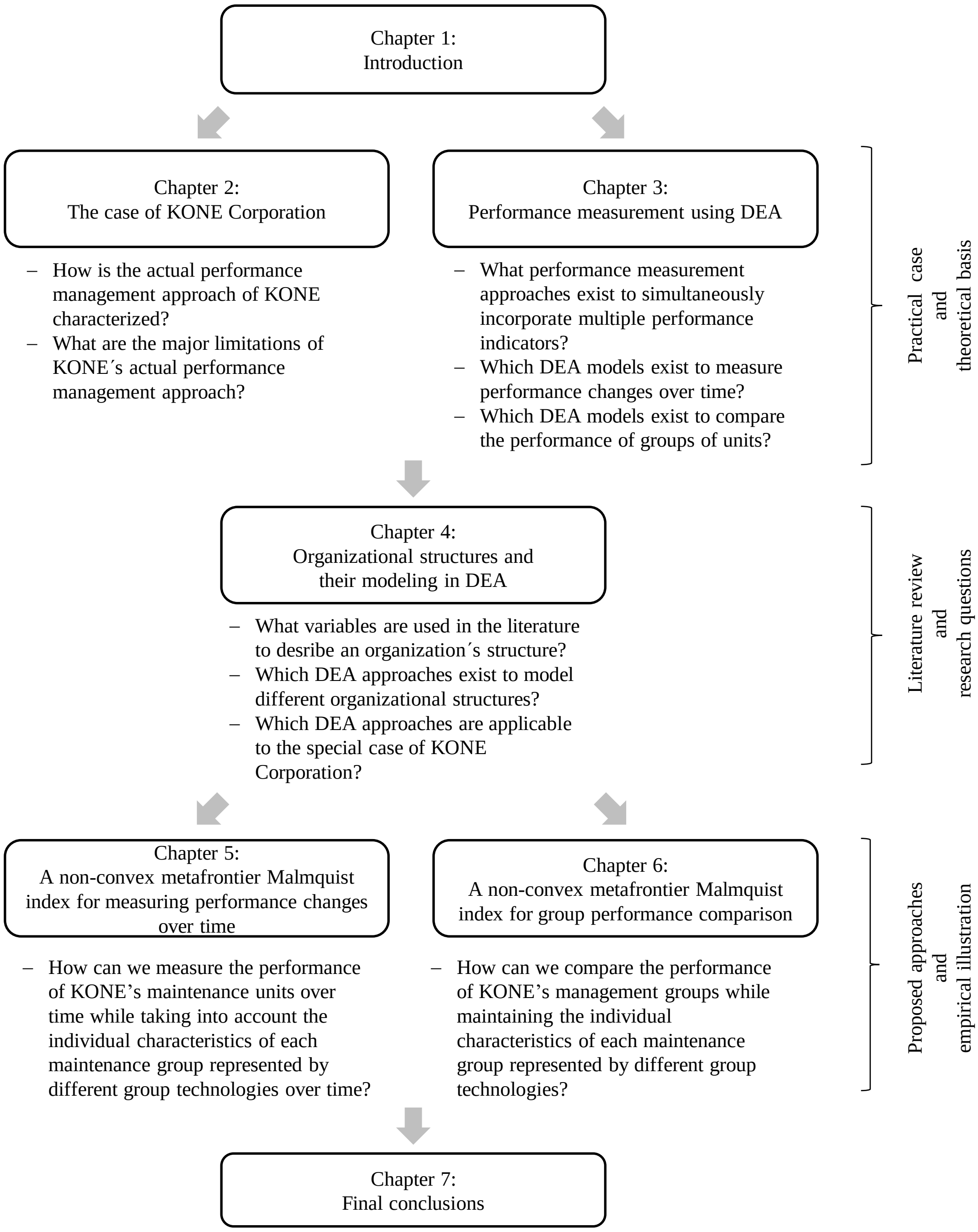
- Chapter 2 provides basic information about the particularities and recent developments in the elevator and escalator industry. Furthermore, a brief description of the focal company KONE Corporation and its local subsidiary KONE GmbH is given. KONE's current framework to manage the performance of its maintenance units and regional groups receives special consideration. This includes a comprehensive discussion of its benefits and drawbacks as well as a summary of requirements for the performance management approach to be proposed.
- Chapter 3 gives a thorough introduction to the theory and mathematical foundations of DEA. Special emphasis is put on the explanation of DEA-based approaches that allow performance changes to be measured over time and compare different groups of operating units. Furthermore, major benefits and limitations of basic DEA models are explained.
- Chapter 4 provides a systematic literature review on how different degrees of centralization are already modeled in the current DEA literature and, in addition, discusses the applicability of the identified approaches to the case of KONE. The chapter ends with the determination of two specific research questions, which are going to be answered subsequently.
- Chapter 5 proposes a non-convex metafrontier-based Malmquist index for measuring performance changes over time, where the panel data comprise groups of units that operate under the influence of different local technologies. The suggested approach overcomes a weakness of the conventional metafrontier Malmquist index, which implicitly neglects that the technology under which each group of units operates can change over time. This negligence may lead to a poor approximation of the metafrontier and accordingly to ambiguous results and flawed managerial conclusions in the case of KONE's maintenance units.
- Chapter 6 extends the respective DEA approach of Camanho and Dyson (2006) for comparing groups of operating units. An alternative index for comparing the performance of the management groups under a centralized management scenario

is proposed. The new approach avoids artificial aggregations of single performance scores (e.g., arithmetic or geometric averages) and is capable of capturing directly the performance of the groups on the basis of their internal abilities in transforming inputs to outputs. The resulting index and its components not only satisfy circularity but also highlight the technological gap in regard to the potential technology available to each group. The proposed approach is also applied to the case of KONE.

- Chapter 7 concludes with a brief summary of the major contributions of this research and describes future research opportunities.

Figure 1.1 illustrates the structure of the thesis.

Figure 1.1: Structure of the thesis



2 The case of KONE Corporation[10]

2.1 Basic company information

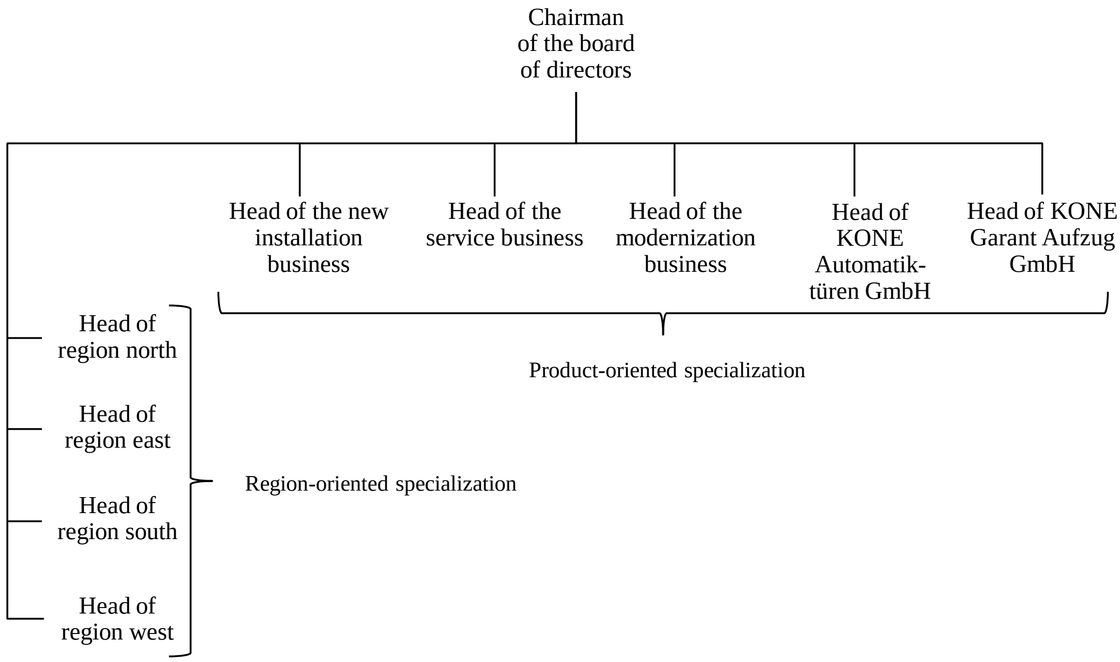
Founded in 1910 in Finland, the international engineering and service company KONE Corporation is recognized today as one of the global leaders in the elevator and escalator industry (see KONE 2018b). Beside elevators and escalators, KONE offers automatic building doors, autowalks and integrated access control systems (see KONE 2018b). The holding company of KONE is organized as a public stock corporation and headquartered in Helsinki (Finland). At the end of the year 2018, KONE generated a total revenue of roughly 8.8 billion Euros (see KONE 2019), which can be separated into three major sources: 55 % of the revenue is generated with the new installation business. The company's maintenance services and performance monitoring technologies contribute 31 % of the yearly revenue. Through the replacement of single equipment parts and the modernization of entire systems, KONE can realize 14 % of its annual revenue (see KONE 2019).

The German subsidiary of KONE is organized as a GmbH (similar to the British business form "Private Limited Company (Ltd)") and headquartered in Hanover (Germany). The organizational structure of KONE GmbH is depicted in Figure 2.1. Since the second management level and also the functional business units (new installation business, service business and modernization business) are divided into geographical regions (North, East, South and West), the organizational structure shows characteristics of a matrix management.

---

[10] This is an updated version of Chapter 3 and 4 of Harms (2016).

Figure 2.1: Excerpt of the organization chart of KONE GmbH



Throughout Germany, KONE established around 70 local maintenance units, which are responsible for the maintenance and repair of elevators and escalators in their respective geographical area. Three to fifteen technicians work in each maintenance unit. In addition, each unit is managed by a local supervisor who primarily coordinates the service operations of his subordinated technicians. In order to oversee its maintenance units in an efficient way, KONE has partitioned them into four distinct managerial groups (e.g., North, East, South and West) with regional headquarters in Hamburg, Berlin, Cologne and Munich, respectively (see Figure 2.1).

Each group is administered by its own regional manager who is responsible for enforcing the overall company targets on the maintenance unit level. The regional managers apply individual management concepts, customized strategies and local procedures which take into account different environmental constraints to run the business according to the demands of their local costumers. The regional managers are also given the authority to reallocate resources between their subordinated units. To this end, collaborations between individual maintenance units are also promoted by regional managers. For instance, maintenance units may request additional support from other group members to handle a task. This is particularly valuable if highly qualified and specialized technicians are required in a specific context.

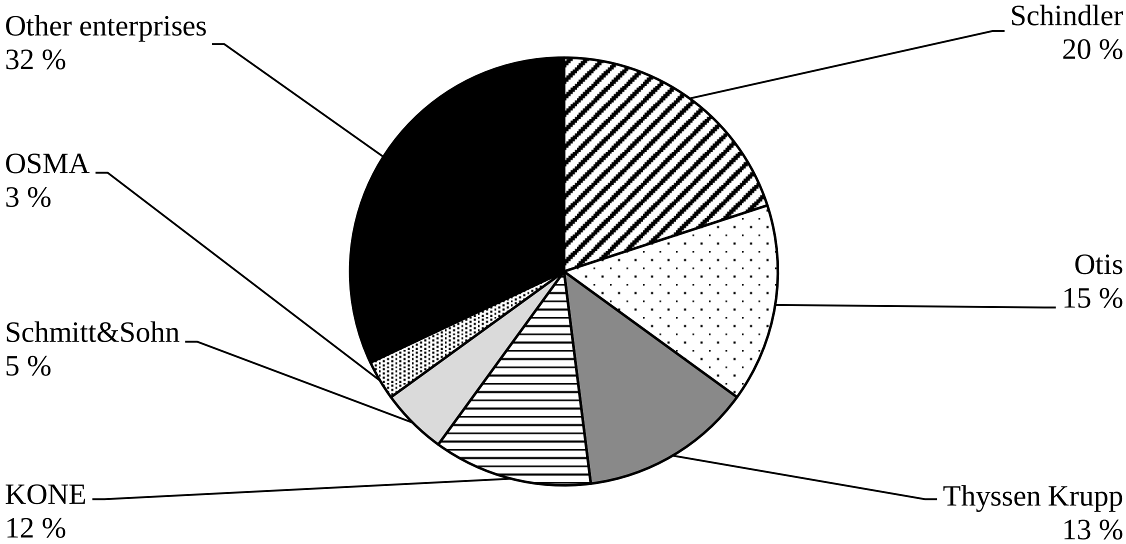
A central back-office team – called the KONE Care Team – supports the different units and regional groups by preparing tenders and undertaking customer discussions. Besides, the KONE Care Team realizes a major part of needed material ordering and compiles the weekly working schedules. The so-called KONE Service Center is the central Call Center of KONE in Germany, and, hence, the central customer contact point. One of the major tasks of the Service Center is to handle emergency calls (e.g., in case of malfunctions or when persons are trapped inside elevators) and, subsequently, inform the supervisors of the local maintenance units about the necessary repair tasks.

2.2 Structure and trends in the elevator and escalator industry

A characteristic feature of the elevator and escalator industry is the existence of three independent and equally important market segments: (1) the production and installation of entire new systems, (2) the modernization of existing systems and (3) the provision of supporting services such as maintenance and repairs (see Dispan 2015, p. 5). The international elevator and escalator market is also characterized by an exceptionally high market-domination of four multinational companies that are, besides KONE, the American company Otis, the Swiss company Schindler and the German company Thyssen Krupp. This group has a combined market-share of approximately 65 % of the international elevator and escalator market and, therefore, is referred to as the “Big-4” (see Dispan 2015, pp. 18-21).

In line with the international market, the German elevator and escalator market is also mainly dominated by the Big-4. For example, in the service sector their combined market-share is roughly 60 %, whereas the remaining market is mainly covered by a few medium-sized businesses operating nationwide (e.g., Schmitt&Sohn, OSMA) as well as several small- and micro-enterprises (see Figure 2.2). The small- and micro-enterprises predominantly focus on market niches or regional customers and especially benefit from their flexibility as well as regional proximity. Besides, these enterprises lack high overhead costs (compared to medium-sized businesses or multinational companies) which allows them to offer much lower service prices (see Dispan 2015, pp. 18-21).

Figure 2.2: Market share distribution in the German elevator and escalator service sector[11]



Despite the dominant market position of the Big-4, the elevator and escalator industry is characterized by fierce price competition causing continuous rationalization and innovation processes in the sector (see VDMA e.V. 2018). However, the price competition has also caused several acquisitions over the past decades. Medium-sized companies in particular were purchased and integrated by one of the Big-4. With such acquisitions, the Big-4 aim at the absorption of technology, know-how, brands and also customer bases. Furthermore, the companies benefit from the takeover of profitable maintenance contracts (see Dispan 2015).

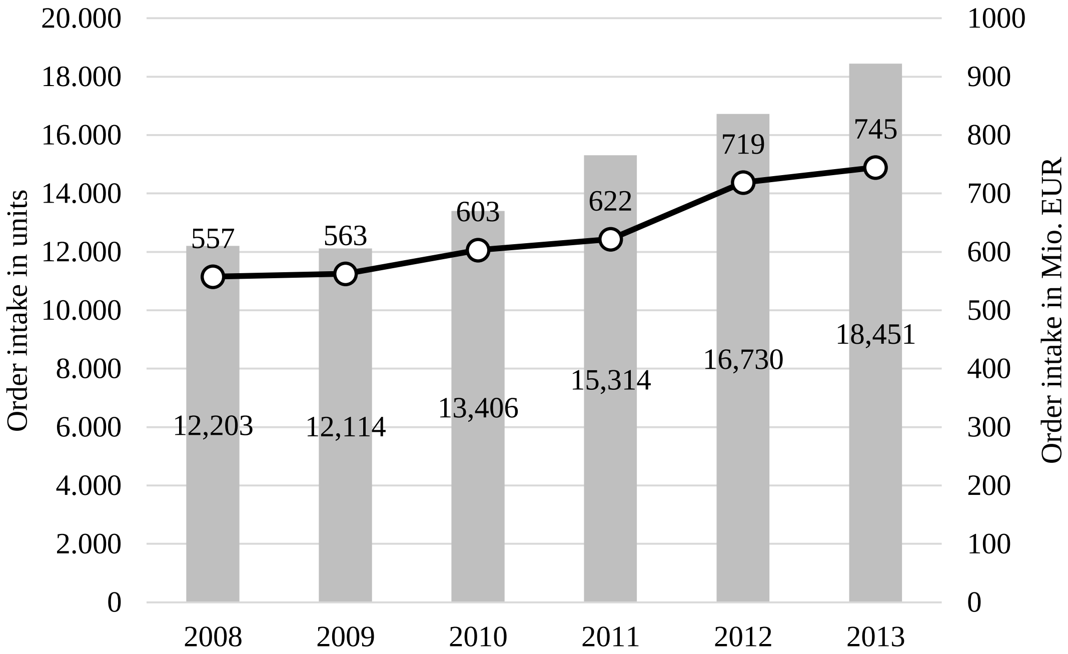
Note that the demand for new elevators and escalators is highly dependent on developments in the construction industry. Corresponding to the recession on the construction market, sales of both elevators and escalators declined (in value and number) from 1998 until 2005 (see Dispan 2007, p. 27). However, after the recovery of the construction business in 2008, only the German *elevator market* continuously increased regarding the value and amount of sold units. By contrast, the order intake on the *escalator market* showed a strong fluctuation and even declined over time (see Dispan 2015, p. 15).

Figures 2.3 and 2.4 depict the indicated development of the order intake in terms of value and quantity on the German elevator and escalator market for the time period between

[11] Source: Dispan (2015, p. 21).

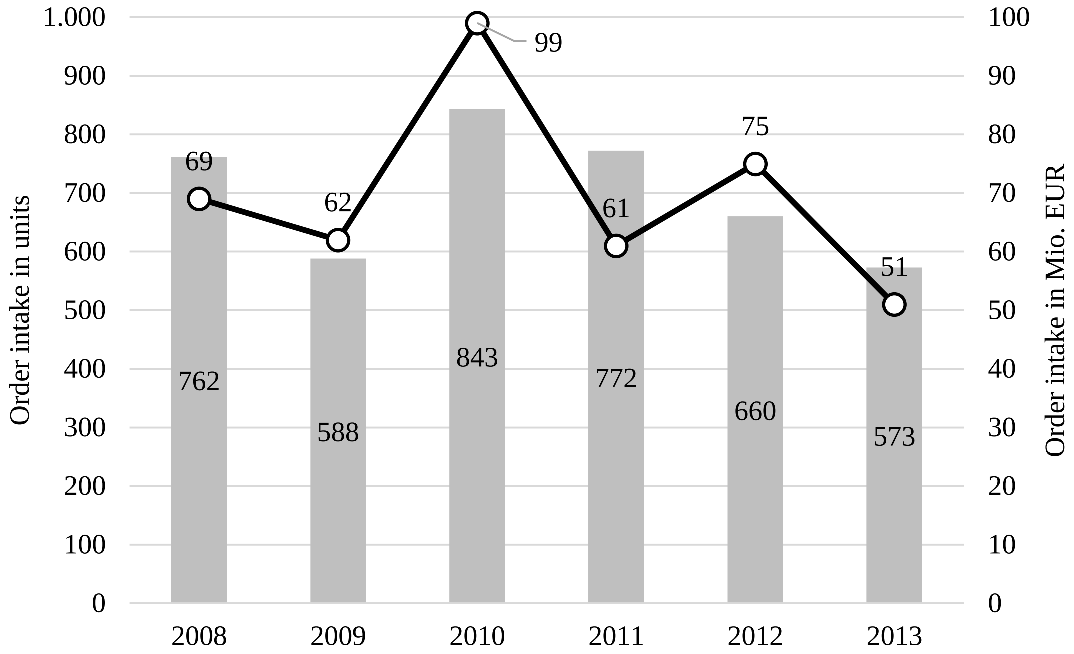
2008 and 2013. Closer inspections of the elevator market development show that – regarding the number of sold units – the sector increased by 8.6 % per year on average, whereas the value of the order intake increased only by 6.0 % per year on average (see Figure 2.3). This means that the average value per elevator has declined since 2008 (from 45,677 Euros per elevator on average in 2008 to 40,370 Euros per elevator on average in 2013). Similar tendencies can be observed for the escalator market: the average price per escalator declined from 90,551 EUR in 2008 to 89,005 EUR in 2013 (see Dispan 2015, pp. 11-12).

Figure 2.3: Order intake development on the German elevator market[12]



[12] Source: Dispan (2015, p. 13).

Figure 2.4: Order intake development on the German escalator market[13]



Such reductions can be interpreted as a direct consequence of the fierce price competition on the (international and German) new sales market. However, other reasons like changed customer tendencies (e.g., towards smaller lift systems) could also have caused such price reductions (see Dispan 2015, pp. 11-12).

Recall that the new installation business depends mainly on the activities of the construction industry and, therefore, follows cyclical fluctuations. Furthermore, the above discussion points to the unstable und problematic price development on the elevator and escalator market. By contrast, maintenance contracts often yield higher profit margins and are mainly independent from other industry developments (see VDMA e.V. 2018). Hence, many companies tend to accept lower (and even unprofitable) prices for new systems if the transactions are combined with profitable long-term service contracts. However, price pressure also increased in the maintenance service segment over the past decades. According to a study of Dispan (2015), the primary causes of this development are:

1. *Professionalization of customers and external consultants*: Large operators have improved their own knowledge regarding elevators and escalators. Furthermore,

[13] Source: Dispan (2015, p. 15)

external consultants with well-founded industrial knowledge enter the market and offer their expertise to department store chains and other key-account customers operating nationwide. These developments improve the position of the operators in negotiations and may have led to lower service prices.

2. *Reduction of maintenance contract duration*: In order to increase flexibility and intensify price pressure, operators of elevators and escalators reduced the average contract terms. In this way, operators often have contract terms of two to five years (at maximum) and can more frequently force price reductions during the next bargaining round.
3. *Negotiations on larger maintenance packages*: Large operators (e.g., department store chains or government agencies) often tender their entire lift system portfolio centrally and publicly. Through this bundling, operators strengthen their bargaining power over service providers and, consequently, can beat down prices even more.

New maintenance concepts and time constraints have caused massive work intensification for technicians simultaneously with the aforementioned developments. Nevertheless, customers are expecting high service quality as well as customer orientation, which requires numerous highly qualified and cost-intensive technicians. These developments have caused increasing rationalization processes in the elevator and escalator industry and forced all market players to continuously improve the performance of their service operations (see Dispan 2015, pp. 50-51).

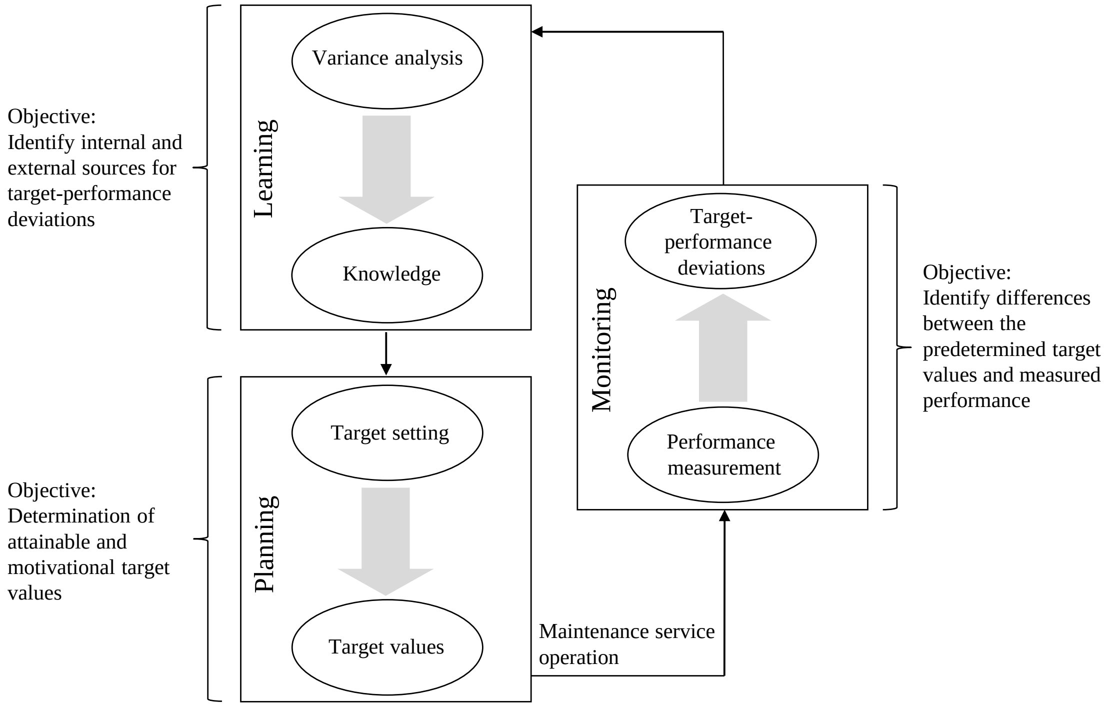
2.3 Traditional performance management

In the process of (service) performance improvement, the application of an appropriate performance management approach is essential. Numerous frameworks and instruments have been proposed by researchers for maintenance-related contexts (see e.g., Groote 1995, Kutucuoglu et al. 2001, Parida and Kumar 2006, Parida and Chattopadhyay 2007 and Muchiri et al. 2011). KONE's approach to controlling the local maintenance units can be separated (based on Ahn 2003) into three distinct steps whereby each step is conducted (at least) once within one business year: (1) planning, (2) monitoring and (3) learning.[14] The different steps and their major objectives are shown in Figure 2.5.

At the end of each fiscal year, KONE's performance management process starts with the *planning* phase. The major objective of this step is to determine motivational targets for each maintenance unit for the forthcoming business year. In order to ensure strategy aligned targets, the specified objective values are derived from KONE's overall business goals, which are centrally determined in the Finnish headquarters for each local subsidiary. Through the application of a top-down budgeting process, the identified targets are substantiated for each managerial region (e.g., North, East, South and West) and each maintenance unit. During this down cascading process, different contextual variables are considered. For example, the size (in terms of employees) and the previous performance of each maintenance unit are taken into account. Besides, the reliability of the lift system inventory, market competition and other potential performance influences are considered. The targets thus derived include financial (e.g., revenue and material cost specifications) and non-financial (i.e., operational) specifications (e.g., number of orders to be completed) for each local maintenance unit on a weekly basis.

[14] See also Ahn (2003), pp. 83-87 for a general explanation of the performance management circle.

Figure 2.5: The performance management process at KONE GmbH[15]



The subsequent *monitoring* phase strongly relies on the correct back-reporting of the technicians of the respective maintenance units. The technicians need to record the required working time, incurred material costs and other relevant information in real time. Based on the entered information and data from other departments (e.g., the KONE Care Team), the enterprise resource planning software (ERP-software) calculates the obtained performance values for each maintenance unit. Subsequently, the computed performance values are compared with the predetermined targets using different system generated reports. This step aims at the identification of deviations between target values and performance values. In compliance with the previously specified target values, the obtained report contains a variety of financial and non-financial performance indicators. Whereas the financial indicators predominantly assess the performance of each unit regarding profit goals, the non-financial measures mainly evaluate the performance of the respective maintenance units in terms of work quality. For instance, the "First-Fixed-Rate" measures whether rework was required (i.e., if system defects could be fixed after the first time the defect occurred). Another possible categorization of KONE's performance indicators

[15] This Figure is based on Ahn (2003, p. 83).

shows that both leading and lagging indicators are included.[16] Whereas the indicator "Service repairs & call out sales" shows how much revenue was generated by the unit in the past (i.e., it is a lagging indicator), the measure "Sales lead" can be considered as an indicator measuring the potential order intake (i.e., it is a leading indicator). In addition, the generated reports contain so-called "Full month run rates" which show whether the predetermined monthly target values will be fulfilled, in the event that the maintenance unit proceeds with its actual level of performance.

The monitoring phase is followed by a *learning* phase. The major objective of this step is to find company-internal and -external reasons for deviations between predefined targets and actual performance values. Therefore, the so-called Productivity Team of KONE discusses the results with the regional management, local supervisors and technicians of each maintenance unit. The outcomes of this phase provide a profound information basis for the target setting of the subsequent business year.

A substantial benefit of KONE's actual performance evaluation process is that the procedure seeks to consider the individual business environment of each maintenance unit during the planning phase. Beside more realistic target values, the consideration of contextual variables promotes the acceptance of the performance management approach at the maintenance unit level. Furthermore, the usage of both financial and non-financial indicators ensures a holistic performance evaluation and allows better predictions of future performance.

However, the approach of KONE also entails some drawbacks worth mentioning here. First of all, it fails to identify the best performing maintenance units. Whereas outperformers can be determined with regard to single indicators, KONE is not able to comprehensively aggregate the different indicators to an overall performance score. So far, such aggregations can only be based on an arbitrary weighting of the different indicators. Besides being highly subjective and prone to flaws, such artificial aggregations usually

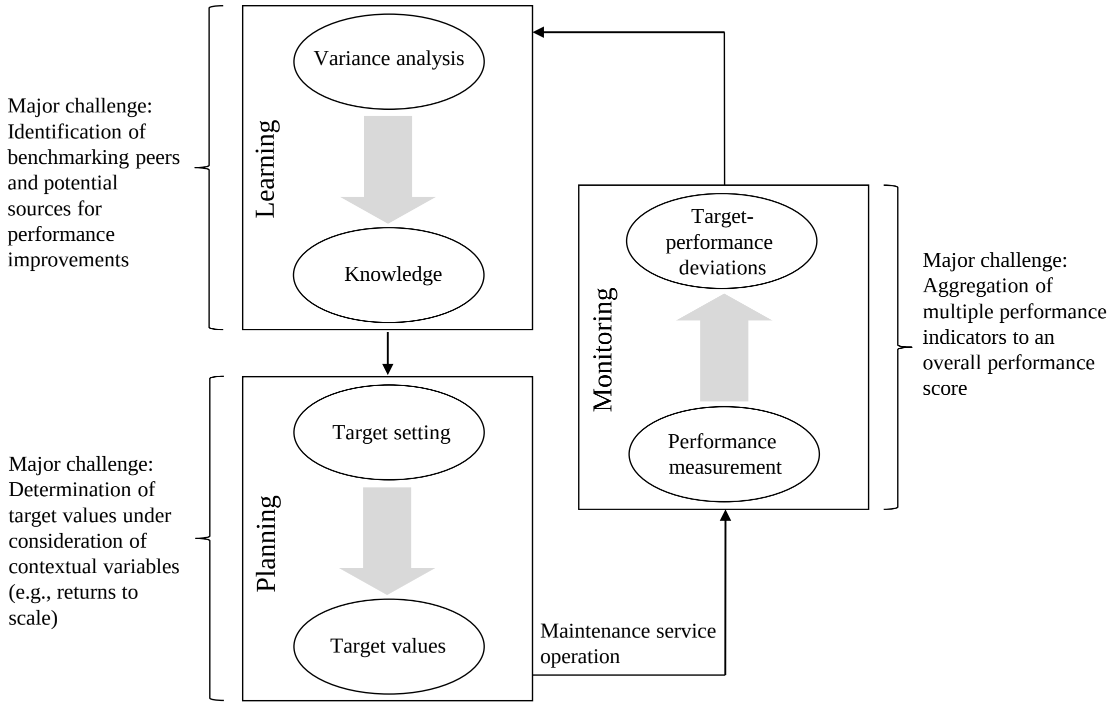
---

[16] Leading indicators provide information on aspects, which are likely to substantially influence the future performance of an organization. By contrast, lagging indicators are usually financial measures which merely allow conclusions about an organization's performance in the past and, therefore, do not provide sufficient insights into the potential future success (see Groote 1995, Kaplan and Norton 1996a, Kaplan and Norton 1996b, Ittner and Larcker 1998 and Bible et al. 2006).

cause numerous discussions with local supervisors and regional managers. In addition, the traditional approach as established by KONE is not able to account for the interdependence between the different variables included in the operational processes of the maintenance units.

The aforementioned deficits also cause other problems. Since the best performing units cannot be identified, many improvement potentials remain undetected. Further problems are caused by the inability of the current approach to detect suitable benchmarking peers which could be used for detailed process analysis. However, such information could possibly supplement the results of the learning phase and, consequently, improve the target setting of the subsequent business year. Figure 2.6 depicts the major challenge of each step of KONE's current performance management approach.

Figure 2.6: Major challenges of the performance management process at KONE[17]



The described drawbacks indicate a possible starting point for further research and, thus, represent the requirements for the yet to be proposed approach. In total, a framework is needed which is able to quantify potentials for improvement and detect benchmarks for

[17] This Figure is based on Ahn (2003, p. 83).

further detail analysis. Thereby, the new method should also be capable of simultaneously including both financial and non-financial variables in the performance assessment (see Groote 1995, p. 4) and aggregating them to an overall performance score. Besides, contextual factors (e.g., returns to scale) and other organizational variables should be included in the determination of performance and target values. To cope with these widely differing challenges at KONE GmbH, Chapter 3 introduces a modern performance measurement approach called DEA.

3 Performance measurement using DEA

3.1 Introduction

Despite its relevance and frequent usage in academic as well as practical circles, the term "performance" is rarely adequately defined (see Tangen 2005). For example, Ahn and Clermont (2018) show that the literature on performance measurement comprises a wide spectrum of different definitions for this term.[18] However, in the majority of cases, the variables "effectiveness" and "efficiency" are considered as "two dimensions of performance".[19]

A definition for efficiency, which is frequently used in production-theoretical contexts, is as follows: "A unit is said to be (pareto-)efficient if it is not possible to raise (lower) any of its output (input) levels without lowering (increasing) at least another one of its output (input) levels and/or without increasing (lowering) at least one of its input (output) levels" (see Thanassoulis 2001, p. 23). Thereby, the term "input" refers to the resources that the respective unit utilizes to produce a bunch of products, here referred to as "outputs".[20]

---

[18] See also Lebas (1995) who gives an overview of different performance measures that are used in maintenance-related contexts.

[19] Ahn and Clermont (2018) refer in their publication to Gilles (2005, p. 20), Cooper et al. (2007, p. 66) and Ozcan (2008, p. 14).

[20] The above definition clearly shows that its foundation relates to production theory. A somehow generalized definition of efficiency based on the foundations of decision theory has been proposed by Ahn (2003) and later successfully adopted by Le (2015) for measuring the performance of German savings banks. However, for the sake of simplicity, the definitions outlined above are consistently used throughout this thesis.

A corresponding definition of effectiveness would be that a unit is said to be effective if it was able to achieve its pre-defined output levels (see Sherman and Zhu 2006, p. 2).[21]

Since effectiveness is based solely on the levels of output and does not account for the consumed input quantities, efficiency is often considered as a more reasonable measure of performance (see Ray et al. 2015, p. 77). Note that the aforementioned definitions implicitly indicate that the perception of performance is highly dependent on the included input and output factors (see Dyckhoff and Ahn 2010, p. 1252), which need to be adapted to the respective research question (see Lebas 1995). For example, one typically evaluates a company's ecological performance based on different indicators than its economic performance.[22]

The above discussions indicate that the concept of performance is rather complex. Furthermore, the corresponding measurement approach should be able to simultaneously incorporate different input and output factors. It is therefore not surprising that traditional performance measurement approaches, which predominantly focus on the examination of single indicators or simple input-output ratios, typically fail to appropriately measure performance (see Thomas et al. 1998, p. 488). This problem is especially dominant when the different indicators cannot be aggregated to an overall performance score on a monetary or other basis.

In these cases, practitioners usually attach specific weights to each indicator and, subsequently, compute an overall performance indicator as the "weighted average" of the various sub-indicators. However, one needs to emphasize that such aggregations based on value judgements are highly subjective and even the application of equal weights (e.g., via the application of the arithmetic average) can be considered as somehow arbitrary (see Zhou et al. 2007). Besides, most of the traditional performance measurement approaches are not able to account for additional factors such as returns to scale or other contextual variables that may affect performance. Correspondingly, these approaches typically yield

[21] A corresponding definition of the "effectiveness" based on the foundations of decision theory has been proposed and applied by Ahn and Neumann (2014).

[22] This is in line with Ahn et al. (2018b) who identified multiple objectives to evaluate the performance of public theatres in Germany. Further evidence is provided by the publication of Ahn and Le (2014) who emphasize the different problems of specifying performance indicators in the case of banks.

inappropriate performance scores which may lead to suboptimal decisions (see Thomas et al. 1998, p. 488).

However, over recent decades two distinct approaches have emerged that allow evaluating the performance of units even in the presence of multiple indicators: the parametric and econometric approach called Stochastic Frontier Analysis (SFA) and the non-parametric approach of DEA. Both SFA and DEA use production frontiers for assessing an entity's performance. Thereby, a production frontier represents the maximum possible output level, which can be produced from a certain amount of input through the application of a production technology (see Ray et al. 2015, p. 76).[23] For the evaluation of a unit's performance, the actual input-output quantities have to be compared with theoretically optimal input-output quantities represented by the respective frontier. However, a major difference between both performance measurement approaches is how they determine this reference frontier.

SFA, which was originally introduced by Aigner and Chu (1968), estimates the unknown production frontier parametrically, i.e. based on different statistical techniques such as linear programming (see Aigner and Chu 1968) or (modified) ordinary least squares (see Richmond 1974). Subsequently, a unit's efficiency can be determined by comparing the optimal output level (represented by the production frontier) with the current output level. A major drawback of SFA is that the estimation of the production frontier is a crucial step, which needs profound knowledge. Hence, this approach is significantly prone to error. Furthermore, SFA cannot handle multiple inputs and multiple outputs simultaneously (see Coelli et al. 2005, p. 241).

By contrast, DEA requires no previous (parametric) specification of the production frontier and is able to simultaneously incorporate multiple inputs and multiple outputs. In DEA, the entities under study are typically referred to as Decision Making Units (DMUs). This implies that each operating unit has control over the process it employs to convert

23 Technology refers to the process by which a unit transforms its inputs into its outputs (see Hackman 2008, p. 1).

its inputs into outputs (see Thanassoulis 2001, pp. 21-22). Therefore, it is only straightforward to claim that these DMUs have an extensive degree of autonomy in making decisions.

In order to measure the performance of DMUs, Charnes et al. (1978) developed a linear programming problem that (endogenously) calculates a set of weights, which maximizes a virtual-output virtual-input ratio of the respective unit under study. That is, the weights, which are attached to the respective input and output factors, are not fixed in advance as in traditional performance measurement approaches. In this way, DEA is able to measure the relative efficiency of DMUs using a minimum set of assumptions about the underlying production technology.

In order to receive a better understanding of DEA, the following subsections are structured as follows: in Section 3.2, the so-called CCR and BCC DEA models are explained. These two approaches are commonly used to determine a DMU's technical and pure technical efficiency. The fundamental benefits and drawbacks of these models are discussed in the same section. Subsequently, Section 3.3 describes two different approaches, which may be used to incorporate characteristics of distinct group technologies into the mathematical models of DEA. In this way, DEA can yield additional information which can supplement a management's decision making. Therefore, these concepts are especially valuable for the particular case of KONE where the maintenance units are grouped according to different management regions. Section 3.4 discusses the application of so-called Malmquist index-based approaches for two different problem settings: first, to compare the performance of single DMUs over time and, second, to compare the performance of entire DMU subsets.

3.2 Basic DEA models

3.2.1 Notations and assumptions

Suppose that there exists a set of $(X_j, Y_j) \in \Re_+^m \in \Re_+^s$ n DMUs ($j = 1,\ldots,n$). Let the respective inputs and outputs of each DMU be expressed by the non-negative and non-zero vectors $X_j = (x_{1j}, x_{2j}, ..., x_{mj})$ and $Y_j = (y_{1j}, y_{2j}, ..., y_{sj})$, respectively. Hence, the technology can be represented by a production possibility set (PPS) or technology set (in the following also abbreviated as "technology") of feasible input-output combinations as follows:

$$PPS = \left\{ (X, Y) \in \Re_+^m \times \Re_+^s \mid X \; can \; produce \; Y \right\}. \tag{3.1}$$

For the sake of simplification, assume that the technology in (3.1) satisfies a minimum set of basic economic assumptions: non-emptiness, free disposability and minimum extrapolation. These distinct assumptions can be mathematically specified as follows:

1. (Non-emptiness). The observed $(X_j, Y_j) \in PPS$, j = 1,…,n.
2. (Free disposability). If $(X, Y) \in PPS$, $X' \geq X$, $Y' \leq Y$, then $(X', Y') \in PPS$.
3. (Minimum extrapolation). PPS is the smallest set which satisfies axioms 1-3.

Taking into account these axioms, the local technologies in (3.1) can be expressed precisely by means of the following technology set:

$$PPS = \left\{ \begin{array}{l} (X \;\; ,Y \;\;) \in \Re_+^m \times \Re_+^s \mid \; x_i \geq \sum_{j=1}^{n} \lambda_j x_{ij}, \; y_r \leq \sum_{j=1}^{n} \lambda_j y_{rj}, \\ \\ \lambda_j \geq 0, \quad j = 1, ..., n \end{array} \right\}. \tag{3.2}$$

3.2.2 Technical efficiency

With respect to the definition of the *PPS* in (3.2), one can measure the distance of a DMU *p* regarding the production frontier using the so-called CCR DEA model. This particular model was introduced by Charnes et al. (1978) and can be mathematically described as follows:

$$
\begin{aligned}
& Eff\ (X_p, Y_p) = min\ \theta_p \\
& s.t. \qquad \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_p x_{ip}, \qquad \forall i \\
& \qquad\qquad \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rp}, \qquad \forall r \\
& \qquad\qquad \lambda_j \geq 0,\ \ \theta_p\ \ free\ in\ sign.
\end{aligned} \tag{3.3}
$$

The constraints of (3.3) require that the activity $(\theta_p X_p, Y_p)$ belongs to the *PPS*, while the objective function seeks the minimum θ_p that reduces the input vector X_p radially to $\theta_p X_p$. Correspondingly, $Eff(X_p, Y_p)$ represents the percentage by which DMU *p* could radially reduce its inputs if everything else is held constant (see Thanassoulis 2001, p. 21). Hence, it can be straightforwardly interpreted as the input-oriented *technical efficiency* score of DMU *p* (see Cooper et al. 2007, pp. 43-44). In order to receive a performance score for each individual DMU, one needs to solve the programming problem (3.3) *n* different times – once for each DMU *j* under observation.

Formula (3.3) has a feasible solution $\theta_p = 1,\ \lambda_p = 1,\ \lambda_j = 0\ \forall\ j \neq p$. Hence, the objective value $Eff(X_p, Y_p)$ is not greater than 1. On the other hand, due to the non-zero assumption for the input and output data (i.e., $(X_j, Y_j) \in \Re_+^m \in \Re_+^s$), constraint 2 forces λ_j to be non-zero for each DMU *j*. Hence, from constraint 1, $Eff(X_p, Y_p)$ must be greater than zero (see Cooper et al. 2007, pp. 43-44) and the performance score is forced to be within the range of 0 and 100 %.

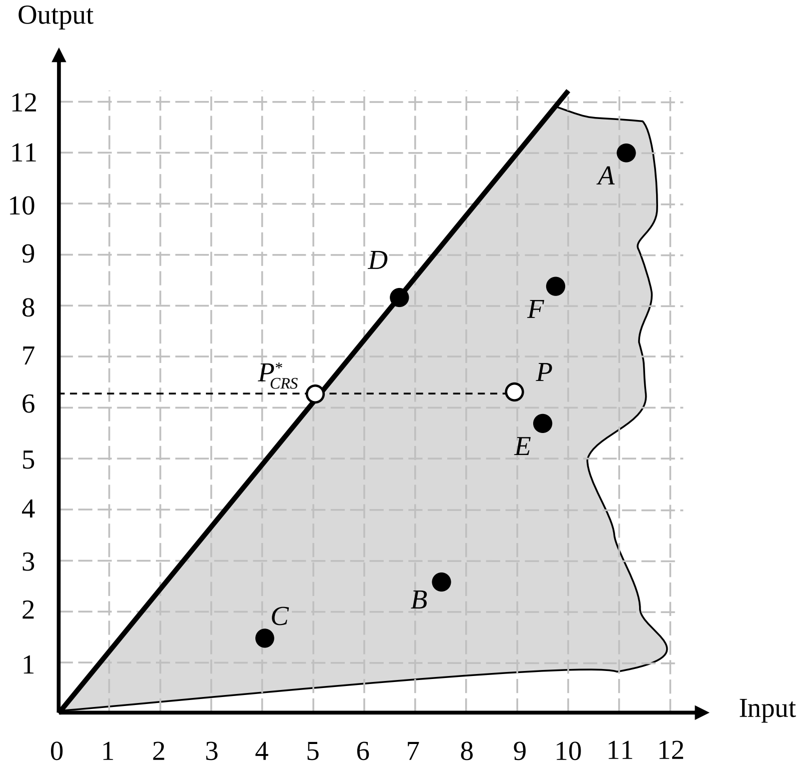
Note that the computed optimal value of $(\theta_p X_p, Y_p)$ is a radial projection of the current performance level of DMU *p* to the facet of the production frontier. The corresponding projection can be considered as a virtual unit that has been obtained by the combination of some “really observed” DMUs. These units are typically referred to as the “reference units” or “benchmarking partners” of DMU *p*. Mathematically, the benchmarking partners for DMU *p* are those units which correspond to positive λ_j. Consequently, a $\lambda_p = 1$ indicates that DMU *p* is operating completely efficiently and is benchmarked against itself.

According to the aforementioned aspects, the optimal value $(\theta_p X_p, Y_p)$ can also be considered as a target value of DMU *p* (see Joro and Korhonen 2015, p. 10) which is exclusively dependent on the observed input-output levels of its reference units. That is, the programming problem in (3.3) explicitly preserves the mix of inputs and outputs of the unit under assessment. Typically, this input-output mix is a direct consequence of the decision autonomy of the respective DMUs and, hence, presents a combination of operating choices (e.g., the relative levels of labor and automation) (see Thanassoulis 2001, p. 78). Consequently, the identified benchmarking partners and received target values implicitly consider the particularities of the DMU under assessment. This should typically improve the practical acceptance of the obtained DEA results.

Even though none of the assumptions mentioned so far define the existing returns to scale, constraints 1-3 implicitly allow that each efficient DMU can increase and decrease its input and output quantities by the same proportion. Hence, the established programming problem of Charnes et al. (1978) is based on the assumption of constant returns to scales (CRS) (see Ray et al. 2015, p. 93). To revisit this implication of the CCR model, examine Figure 3.1. It illustrates the *PPS* of the CCR model for the single-input single-output case as a shaded area. The bold line depicts the corresponding production frontier. The projection of DMU *P* regarding this production frontier is represented by point P^*_{CRS}. As a consequence, the distance between point *P* and P^*_{CRS} represents the input-oriented efficiency of DMU *P* which is determined through the solution of (3.3).

One can also conclude from Figure 3.1 that only one DMU (e.g., DMU *D*) lies exactly on the production frontier, which is correspondingly considered as the only efficient unit. The remaining DMUs can theoretically reduce their input quantities without deteriorating their current output level. Hence, it is straightforward to classify these DMUs as inefficient.

Figure 3.1: The DEA-based production frontier of a CRS-technology



The mathematical representation given by (3.3) is sometimes referred to as the "envelopment form" (or "Farrell form") of the CCR DEA model. This envelopment form can be transformed using the so-called Charnes-Cooper-Transformation (see Charnes and Cooper 1962) into the following "multiplier form" which provides additional managerial information:

$$
\begin{aligned}
& Eff(X_p, Y_p)^{-1} = max\ \phi_p \\
s.t.\quad & \phi_p = \sum_{r=1}^{s} \mu_r y_{rp} \\
& \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} \leq 0, \qquad \forall j \\
& \sum_{i=1}^{m} \nu_i x_{ip} = 1 \\
& \mu_r, \nu_i \geq \varepsilon, \qquad \forall r, \forall i \\
& \varepsilon > 0,\ \ \phi_p \ \ free\ in\ sign.
\end{aligned}
\tag{3.4}
$$

In (3.4), ν_i and μ_r denote the weights which are attached to input *i* and output *r*. The constraint $\sum_{i=1}^{m} \nu_i x_{ip} = 1$ prevents an infinite number of solutions being possible (ν_i and μ_r would be unlimited otherwise). ε is a non-Archimedean element smaller than any positive real number. The constraint $\varepsilon \geq 0$ guarantees that the calculated solutions of ν_i and μ_r are positive and, hence, none of the variables is neglected in the evaluation process (see Cooper et al. 2011b, pp. 9-11).[24] Note that in the context of DEA-based input-oriented efficiency measurement, the reciprocal of the optimal value received from the multiplier model is identical to the received efficiency score from (3.3).

After the solution of (3.4), one obtains for each DMU an individual set of weights, which is denoted as ν_i^* and μ_r^* in the following. This set of weights puts the performance of the respective DMU under evaluation in the "best possible light". This is why ν_i^* and μ_r^* are sometimes referred to as the "most favorable set of weights". From this set of weights, one can conclude not only "which performance criteria" contribute to the performance score of DMU *p* but also to "what extent" they do so. Hence, these values indicate the relative importance of each item for the respective DMU under evaluation (see Cooper et al. 2007, p. 25).

3.2.3 Pure technical efficiency

Since the basic CCR DEA model is based on the CRS specification, it is implicitly assumed that every DMU has the optimal scale size. However, the particular business environment may cause scale inefficiencies, i.e. inefficiencies solely caused by a non-optimal size of the DMU. For these instances, the application of the CCR DEA model may cause inappropriate DEA results which may not guarantee optimal decision making (see Coelli et al. 2005, p. 172). Against this background, Banker et al. (1984) proposed a DEA

[24] Note that, here, the multiplier form as introduced by Charnes et al. (1979) is given. In the original version published by Charnes et al. (1978), ν_i and μ_r have been allowed to take zero-values. However, as this implicitly means that these factors might not be considered within the evaluation process, Charnes et al. (1979) suggested a slightly modified version which uses the non-Archimedian element ε and, therefore, prevents some factors from being neglected in the evaluation of a certain DMU.

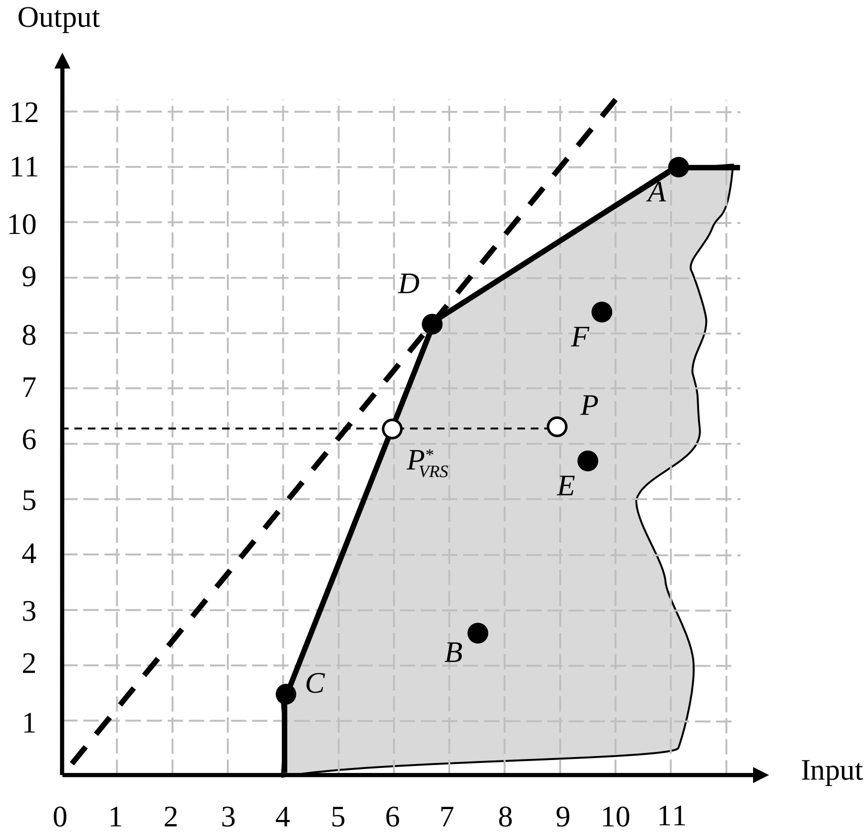
model which assumes variable returns to scale (VRS) and is typically referred to as the BCC DEA model. This model enables the calculation of efficiency scores which are not distorted by scale inefficiencies (see Coelli et al. 2005, p. 172). Hence, the corresponding efficiency score can be considered as the *pure technical efficiency* of a DMU (see Cooper et al. 2007, p. 141). The input-oriented envelopment form of the BCC model is outlined below:

$$\begin{aligned}
& Eff\,(X_p, Y_p) = min\ \theta_p \\
& s.t. \qquad \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_p x_{ip}, \qquad \forall i \\
& \qquad\qquad \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{rp}, \qquad \forall r \\
& \qquad\qquad \sum_{j=1}^{n} \lambda_j = 1 \\
& \qquad\qquad \lambda_j \geq 0,\ \ \theta_p\ \ free\ in\ sign.
\end{aligned} \tag{3.5}$$

In terms of model representation, constraints 1-3 of (3.3) are identical to constraints 1, 2 and 4 of (3.5), respectively. The interpretation and the corresponding managerial implications remain unchanged; hence, they are not repeated here. Constraint 3 of (3.5) restricts the construction of virtual reference units by enabling only convex combinations of real observed units. In other words, this constraint is the mathematical expression of the VRS specification. Assuming VRS automatically ensures that inefficient DMUs are only benchmarked against units with a comparable scale size. By contrast, in the CRS model, units may be compared to DMUs, which are substantially larger or smaller (see Coelli et al. 2005, pp. 172-173).

Figure 3.2 illustrates the underlying *PPS* of (3.5) for the previously discussed single-input single-output case. Again, the corresponding *PPS* is shown as a shaded area. The VRS production frontier is depicted by a bold line and the CRS production frontier is represented by a dashed line.

Figure 3.2: The DEA-based production frontier of a VRS-technology



One can conclude from Figure 3.2 that the *PPS* for CRS models is larger than that for VRS counterparts, i.e. $PPS_{VRS} \subseteq PPS_{CRS}$. This finding can be, initially, counterintuitive; the CRS model is less constrained than the VRS model because the mathematical representation in (3.5) has one additional constraint. However, the more constrained the model, the lower the chance of a single DMU being declared inefficient (see Charnes et al. 2013, p. 71). As a direct consequence, the pure technical efficiency scores of the BCC model are equal to or greater than the corresponding technical efficiency values of the CCR model.

This implication can also be straightforwardly derived from Figure 3.2. For example, the distance of DMU *P* to the VRS production frontier (i.e., P- P^*_{VRS}) is smaller than the associated distance to the CRS frontier (i.e., P- P^*_{CRS}). In line with these distance gaps, the respective researcher can identify more improvement potentials for DMU *P* under CRS than under VRS. This corresponds to a larger efficiency score under VRS compared to CRS. Furthermore, Figure 3.1 shows that the number of efficient DMUs increased: under the assumption of a VRS technology, DMUs *A*, *C* and *D* are declared efficient. In contrast to that, only one efficient DMU (e.g., DMU *D*) appears under the assumption of the CRS technology.

Note that it is also possible to enforce other returns to scale assumptions to model (3.5) via a simple mathematical modification of the third constraint. For example, to compute the pure technical efficiency of a DMU under non-increasing returns to scale or non-decreasing returns to scale, constraint 3 can be replaced by $\sum_{j=1}^{n}\lambda_j \leq 1$ or $\sum_{j=1}^{n}\lambda_j \geq 1$, respectively (see Banker et al. 1984).

Just like the CCR model, the envelopment form of the BCC model can also be transformed via the application of duality theory into a corresponding multiplier form (see Banker et al. 1984). The following formula (3.6) shows the respective linear programming problem:

$$\begin{aligned}
& Eff(X_p, Y_p)^{-1} = max\ \phi_p \\
& s.t. \quad \phi_p = \sum_{r=1}^{s} \mu_j y_{rp} + u \\
& \qquad \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} + u \leq 0, \qquad \forall j \\
& \qquad \sum_{i=1}^{m} \nu_i x_{ip} = 1 \\
& \qquad \mu_r, \nu_i \geq \varepsilon, \qquad \forall r, \forall i \\
& \qquad \varepsilon > 0,\ \ \phi_p, u\ \ free\ in\ sign.
\end{aligned} \tag{3.6}$$

The free variable *u* is the dual variable associated with constraint 3 in the envelopment form and, hence, does not appear in the CCR model. The remaining constraints and corresponding economic interpretations are identical to what has already been explained for the multiplier form of the CCR model. Therefore, they shall not be explained here again. However, it should be emphasized that even (3.6) implies that each DMU can choose an individual set of weights which puts its respective performance in the best possible light.

This can be concluded from the DMU-specific weight variables ν_i and μ_r which are only bounded by the non-negativity constraint (i.e. $\mu_r, \nu_i \geq \varepsilon \; \forall r, \forall i$).

So far, it has been assumed that all inputs need to be reduced and all outputs need to be increased. However, one may also encounter situations where both desirable (good) and undesirable (bad) output and input factors are present. For instance, in a paper mill production plant undesirable outputs such as pollutants (e.g., biochemical oxygen demand, suspended solids, particulates and sulfur oxides) may occur (see Seiford and Zhu 2002). These particular input and output sets contradict the assumptions of traditional DEA models which consistently require that an increase in outputs or decrease in inputs is desirable in terms of efficiency improvement. As a direct consequence, these special cases cannot be solved with the aforementioned approaches (see Ali and Seiford 1990).

However, the respective researcher may take advantage of the so-called translation invariance property of the BCC model (see Ali and Seiford 1990). Therefore, each desirable input factor or undesirable output factor needs to be multiplied by -1 and then translated with a proper vector to let all negative factors be positive. Subsequently, these translated data sets can be applied in the basic BCC model to compute the respective performance score for each DMU (see Seiford and Zhu 2002). Note that this displacement of the data does not alter the VRS production frontier and, correspondingly, the classification of DMUs as inefficient or efficient. This is a substantial advantage which is not shared by the CCR model. However, even in the case of the BCC model, the performance scores (i.e., the received objective function values) obtained for the inefficient DMUs are different when the data is translated (see Ali and Seiford 1990, p. 405).

3.2.4 Benefits

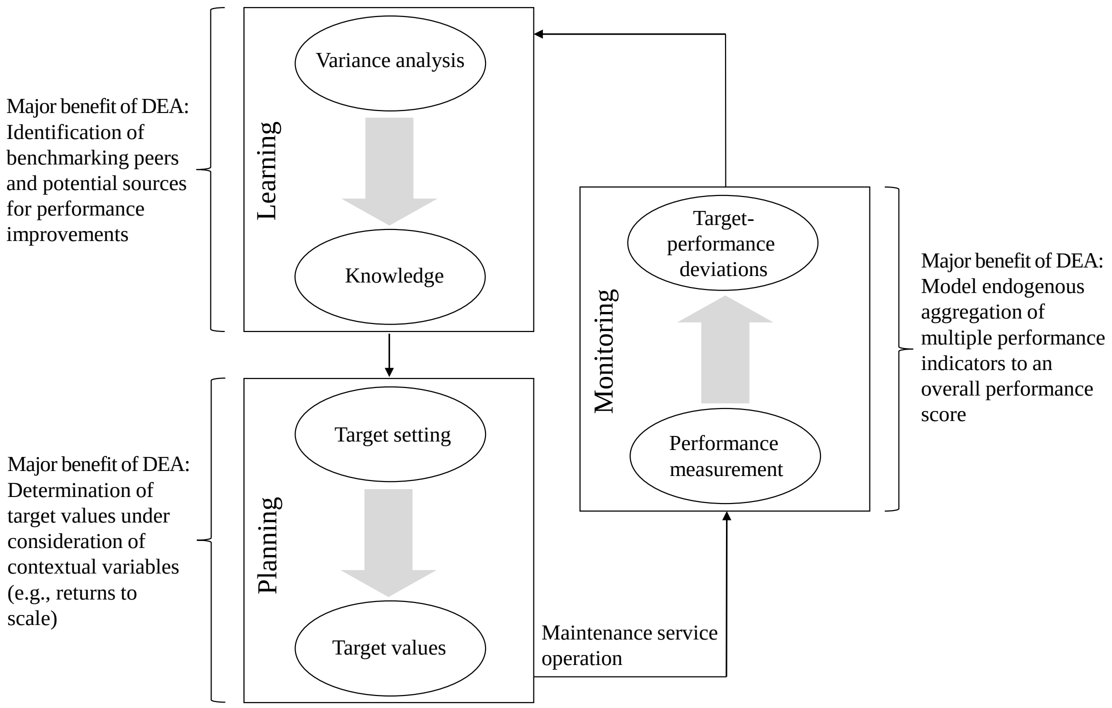
Compared to traditional approaches, DEA entails some important benefits that make this method attractive to several practical situations. Some of these benefits have already been mentioned in the course of this chapter. However, to underpin the usefulness of DEA for the practical case of KONE Corporation and highlight how DEA can comprehensively facilitate the different phases of the performance management circle of Ahn (2003), three noteworthy benefits are discussed in more detail below. Furthermore, a graphical overview of the different benefits – classified according to the performance management circle of Ahn (2003) – is presented in Figure 3.3.

1. *Target values*: Based on radial projections onto the reference technology, basic DEA models yield an efficiency score for each unit that gives information on the respective distance from the production frontier (see Coelli et al. 2005). One important by-product of this assessment is a set of individual target values (for each input and output variable) that would render the DMU under consideration relatively efficient. Within the corresponding mathematical computations, DEA is able to incorporate important production theoretical considerations such as returns to scale or the restricted controllability of inputs and outputs (see Thanassoulis and Dyson 1992).[25] These characteristic features of DEA are not only a clear difference to traditional approaches published in the current business science literature, but can lead to more realistic and acceptable target values for the different DMUs.
2. *Model endogenous weighting*: In order to receive a single overall performance measure, practitioners usually attach a set of fixed weights to the different indicators. However, since the weights are determined by management representatives or other experts, they are not only highly subjective but also prone to misspecifications caused by behavioral biases. By contrast, basic DEA approaches derive the input and output weights directly from the applied data set. That is, numerous a priori assumptions and computations involved in fixed weight choices are avoided (see Cooper et al. 2006). This method of deriving weights is usually referred to as "model endogenous weighting" meaning that the different indicator weights are a major *outcome* of the analysis. An associated advantage of the model endogenous weighting of basic DEA models is that discussions with a DMUs' management about an alternative weighting scheme become obsolete, since any other set of weights would not improve the resulting performance scores (see Ahn 2014).

[25] See Thanassoulis and Dyson (1992) for a thorough discussion on target setting using DEA.

3. *Identification of potential sources for performance improvement*: In order to compute the respective performance scores, the basic DEA models radially project the actual input-output-levels of the DMU under evaluation to the production frontier, which is constructed by entities operating in a comparable field or industry. Based on these projections, DEA can also provide information about a set of efficient "benchmarking partners" for each inefficient DMU. These efficient peers have a similar (or even identical) mix of input-output levels to that of the DMU under evaluation (but with a lower input or higher output level). This usually makes these efficient peers suitable for detail process analysis or role models to improve the performance of the inefficient DMUs (see Thanassoulis 2001).

Figure 3.3: Benefits of basic DEA models[26]



3.2.5 Limitations

Since the appearance of DEA, this methodology has been applied to various practical situations, revealing different limitations. The major limitations of basic DEA models are described below. They have been categorized into four different streams:

---

[26] This Figure is based on Ahn (2003, p. 83).

1. *Inappropriate specification of the reference technology*: Because of the model-endogenous determination of the input and output weights, DEA is usually considered as a performance measurement approach which substantially reduces the arbitrariness and subjectivity of the evaluation process. As highlighted by Dyckhoff and Ahn (2010), some authors even argue that DEA is a completely objective performance measurement tool (see e.g., Bouyssou 1999). However, one needs to recall that evaluations based on DEA are highly dependent on the applied reference technology. Again, this is specified by the respective researcher based on professional judgement (see Dyckhoff and Ahn 2010, p. 1252, Afsharian et al. 2016, pp. 1892-1893). For example, the researcher needs to select inputs and outputs, which have to appropriately reflect the underlying transformation process of the operating units. This is a very crucial step, as, e.g., the ongoing discussions regarding the selection of inputs and outputs in the context of bank branch efficiency evaluations show (see e.g., Berger and Humphrey 1997, Fethi and Pasiouras 2010, Paradi et al. 2011, Ahn and Le 2015, Ahn and Le 2016). Furthermore, the determination of the returns to scale as increasing, decreasing, constant, variable etc. may be another step where the definition of the reference technology is prone to misspecifications. There are various other examples where professional judgement needs to be applied to appropriately define the reference technology (e.g., free disposability, convexity). Altogether, the subjectivity of the respective researcher may cause an unrealistic estimation of production frontier and, in the end, may yield misleading performance scores and interpretations.
2. *Low discrimination power of performance scores*: The basic DEA models often rate a large proportion of units as efficient and, therefore, do not allow a sufficient discrimination between the performances of these DMUs. However, the discrimination between inefficient DMUs may also be problematic, especially when the number of units is relatively small compared to the total number of incorporated inputs and outputs (see Andersen and Petersen 1993, p. 1354). This is because the introduction of a large number of factors tends to shift the compared units towards the production frontier, resulting in a relatively large proportion of units with high efficiency scores (see Golany and Roll 1989, p. 240). Because of this poor discrimination power, DEA may not be appropriate for cases where the decision

maker has to determine the best unit or rank DMUs according to their received performance scores (see Karsak and Ahiska 2005, p. 1543).

3. *Inappropriate weighting schemes*: The total weight flexibility of DEA is one of the most appealing aspects of this methodology (see Cooper et al. 2011a, p. 95). As the mathematical models in DEA are run separately for each DMU, the set of weights is usually different for the various DMUs (see Roll et al. 1991, p. 2). In some cases, the different factors may also receive a negligible weight, meaning that these factors are in fact ignored in the efficiency assessment (see Roll et al. 1991, p. 3). However, many authors have criticized such extreme weighting schemes, especially if the performance regarding certain input or output factors cannot be ignored in practice. For example, when assessing different university departments with the two outputs “number of graduated master students” and “number of graduated bachelor students”, it might be hard to justify that a university department attaches a negligible weight to one of the two outputs when governmental regulations demand the education of both student types. As a result, the relative efficiency of a DMU may be flawed and not appropriately reflect its real performance. In specific cases, this means that a certain DMU may be evaluated as relatively efficient merely because its ratio for a certain, possibly irrelevant, input-output combination is the highest when compared to the remaining DMUs in the data set (see Dyson and Thanassoulis 1988, p. 564). Besides, the flexibility often leads to unreasonable results in the sense that the attached weights are frequently inconsistent with expert knowledge or previous expectations regarding the DMUs’ transformation process (see Cooper et al. 2011a, p. 95). Note that flexibility in the selection of weights can be substantially important for its practical applicability when DEA is combined with an incentive system. When the unit under assessment is not rated as 100 % efficient, this implies that the remaining DMUs are more productive even when the weights of all units are set to maximize the score of the unit assessed (see Khalili et al. 2010). Therefore, the inefficient unit cannot argue that its received performance score would be better if a different set of weights is applied. However, it may also be difficult to reason why widely differing weights are attached to the same factor only because different units are

evaluated (see Roll et al. 1991, p. 3). This in turn may contradict the comparability of the DMUs and again lead to lower practical acceptance.

4. *Inappropriate projections and targets*: In many real-world applications, DMUs are only rarely restricted by regulations or policies and, therefore, allowed to concentrate their efforts on a few factors to improve their performance. This flexibility builds the fundamental idea behind basic DEA models which is why such specialized DMUs may also obtain high efficiency scores (see Ahn et al. 2012, pp. 417-418, Dyckhoff et al. 2013, pp. 40-41). However, there are also cases where the performance of DMUs needs to be measured in line with an overall policy or directive. For example, national governments may apply regulation mechanisms to have considerable influence on the decisions, strategies, objectives and resources of operating entities in network industries (e.g., electricity, natural gas, water supply and telecommunication) (see Afsharian et al. 2019a). A concentration of DMUs only on some factors may contradict this regulatory mechanism, which is why the received projections and targets of basic DEA models may be inappropriate for such regulatory regimes. In order to better reflect the preferences and corresponding expectations of the respective evaluator, it may be straightforward to measure a DMU's efficiency according to a previously specified direction of measurement which is consistent with the corporate strategy and overall goals of the evaluator (see Afsharian and Ahn 2014).

3.3 Metafrontier-based performance measurement

3.3.1 Background

Note that the aforementioned DEA models are based on a bunch of different homogeneity assumptions about the underlying set of DMUs. A first assumption is that all DMUs make use of similar resources (e.g., employed personnel, auxiliary materials etc.) so that a common set of input factors can be specified. A second assumption implies that the units transform the specified inputs into comparable products and/or services which is expressed in the common set of outputs. Finally, there is an implicit assumption that the

DMUs operate in similar business environments, since the external environment generally impacts the overall performance of units (see Dyson et al. 2001, p. 247).

Note that the aforementioned assumptions about the set of DMUs are satisfied only in very rare instances. In most situations, the DMUs may operate in different industries, regions and/or countries and correspondingly face different customers, competitors and production opportunities. Because of such differences in the units' environment, researchers usually tend to estimate separate production frontiers for different groups of DMUs (see O'Donnell et al. 2008). For example, separate frontiers have been applied for evaluating banks in Greece (see Vassiloglou and Giokas 1990), the United States of America (see Yue 1992) and Canada (see Cook et al. 2000).

After estimating a local production frontier, it is common and straightforward to measure the relative within-group efficiency of DMUs (e.g., banks of Greece). However, there is often considerable interest to compare the performance of units across the different groups (e.g., comparing the performance of banks from Greece with the performance of banks from Canada). Unfortunately, such comparisons are only meaningful when the respective units are evaluated regarding the same production frontier (see O'Donnell et al. 2008, p. 232).

In line with this idea, Hayami (1969) as well as Hayami and Ruttan (1970) proposed a so-called metafrontier for evaluating all DMUs regardless of their respective group membership. This metafrontier can be considered as a global benchmark technology that is the reference production frontier for all units under assessment. This concept has been extended by Battese and Rao (2002), Battese et al. (2004) as well as O'Donnell et al. (2008) for its application to SFA and DEA.

In DEA, the metafrontier is formed as the all-encompassing frontier that envelops all group frontiers (see O'Donnell et al. 2008, pp. 231-232). Its estimation can be conducted in two different manners whereby each technique relies on a different set of fundamental assumptions about the characteristics of the group technologies. The first technique specifies the metatechnology as a convex combination of the individual group technologies. The scientific literature refers to this type of metafrontier as a *convex metatechnology*. The second estimation technique rejects that convexifications between individual group technologies is attainable for the different DMUs. Therefore, the corresponding metafrontier is usually called a *non-convex metatechnology*. Both possibilities of defining the metatechnology are described in more detail in Subsections 3.3.2 and 3.3.3 below, respectively.

3.3.2 Convex metatechnology

Suppose that there exists a panel of *n* DMUs which can be partitioned into *G* (*G*>1) distinct groups. Let each group *g* (*g*=1,…,*G*) include δ_g DMUs $(X_j^g, Y_j^g) \in \Re_+^m \in \Re_+^s$ (j=1,…, δ_g), where $X_j^g = (x_{1j}^g, x_{2j}^g, \ldots, x_{mj}^g)$ and $Y_j^g = (y_{1j}^g, y_{2j}^g, \ldots, y_{sj}^g)$ are non-negative and non-zero vectors of inputs and outputs, respectively. Following O'Donnell et al. (2008), it is also assumed that all DMUs in each group *g* operate under the same technology, resulting from, e.g., the same resource, regulatory or other environmental constraints. Hence, each local technology of group *g* can be represented by a *PPS* of feasible input-output combinations as follows:

$$PPS^g = \left\{ (X^g, Y^g) \in \Re_+^m \times \Re_+^s \,\middle|\, X^g \text{ can produce } Y^g \right\}. \tag{3.7}$$

Suppose that each local technology in (3.7) satisfies the same axioms as defined in Section 3.2.1 (i.e., non-emptiness, free disposability and minimum extrapolation). In this case, the local technologies can be expressed precisely by means of the following technology sets:

$$PPS^g = \left\{ \begin{array}{l} (X^g, Y^g) \in \Re_+^m \times \Re_+^s \,\middle|\, x_i^g \geq \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g,\ y_r^g \leq \sum_{j=1}^{\delta_g} \lambda_j^g y_{rj}^g, \\ \lambda_j^g \geq 0, \quad j = 1, \ldots, \delta_g \end{array} \right\}. \tag{3.8}$$

With respect to the definition of PPS^g in (3.8), one can measure the within-group efficiency of a DMU against the frontier of a particular group *g* (*g* = 1,…,*G*) using the basic CCR DEA model given by (3.3). Moreover, the convex metatechnology can be defined as (see e.g., O'Donnell et al. 2008; Huang et al. 2013)

$$PPS_C^M = \bigcup_{g=1}^{G} PPS^g \tag{3.9}$$

whereby PPS_C^M aggregates all group technologies and the subscript "C" indicates that this aggregation is based on convex combinations. Hence, PPS_C^M can be precisely expressed by means of the following technology set:

$$PPS_C^M = \left\{ \begin{array}{l} (X^g, Y^g) \in \Re_+^m \times \Re_+^s \mid \; x_i^g \geq \sum_{g=1}^{G} \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g, \; y_r^g \leq \sum_{g=1}^{G} \sum_{j=1}^{\delta_g} \lambda_j^g y_{rj}^g, \\ \lambda_j^g \geq 0, \quad j = 1, \ldots, \delta_g \end{array} \right\}. \tag{3.10}$$

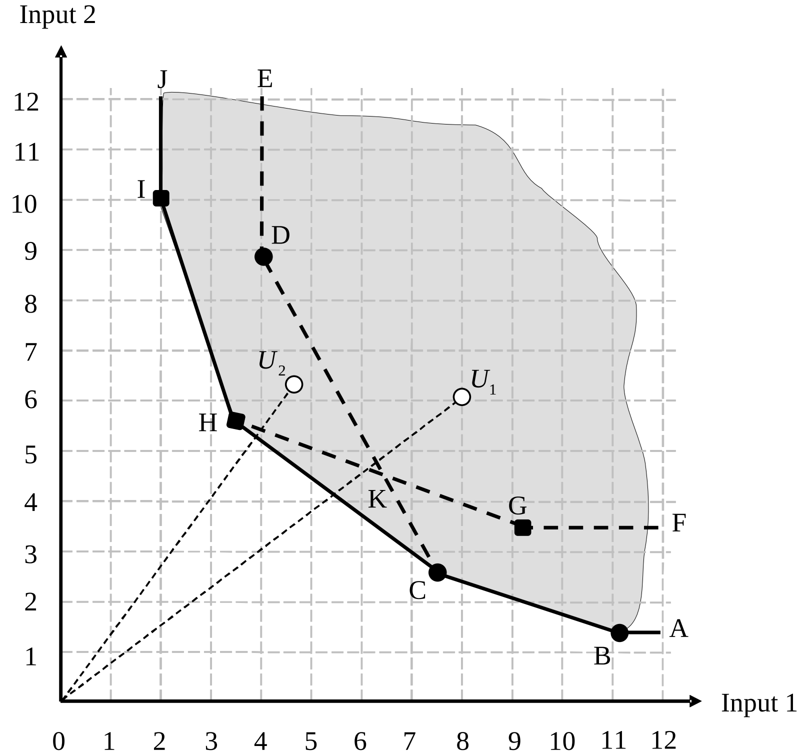
In reference to this global benchmark technology, the metafrontier (technical) efficiency of a DMU *p*, denoted in the following as $Eff_C^M(X_p^g, Y_p^g)$, can be computed using the following mathematical programming problem:

$$\begin{array}{ll} Eff_C^M(X_p^g, Y_p^g) = min \; \theta_p^C & \\ s.t. \quad \sum_{g=1}^{G} \sum_{j=1}^{\delta_g} \lambda_j x_{ij}^g \leq \theta_p^C x_{ip}, & \forall i \\ \qquad \sum_{g=1}^{G} \sum_{j=1}^{\delta_g} \lambda_j y_{rj}^g \geq y_{rp}, & \forall r \\ \qquad \lambda_j \geq 0, \;\; \theta_p \; free \; in \; sign. & \end{array} \tag{3.11}$$

Note that (3.11) shows substantial similarities with the linear programming problem given by (3.3) which has been formulated to compute the within-group efficiency. For example, both approaches are based on the CRS assumption. However, constraints 1-2 of (3.11) aggregate all observations of all *G* groups to form the benchmark technology and, hence, incorporates additional information in building the evaluation facet.

Revisit this idea and its different practical implications based on Figure 3.4. Here, it is supposed that there exist two group technologies T^1 and T^2 with two inputs and a single output. As the single output is assumed to be the same for all DMUs, it is not shown. Hence, the group technologies are depicted as the areas bounded by ABCDE and FGHIJ, respectively. Furthermore, the convex metatechnology is represented by ABCHIJ.

Figure 3.4: Convex estimation of a contemporaneous technology set



Suppose the respective researcher seeks to evaluate the efficiency of two distinct units, depicted as U_1 and U_2 in Figure 3.4, whereby each of the units belongs to a different group (e.g, group 1 and 2). Hence, the efficiency of U_1 regarding its own group frontier (i.e., the border shown by ABCDE) can be computed via the application of model (3.3). The associated within-group efficiency index is denoted in the following as $Eff(X_{U1}^{g=1}, Y_{U1}^{g=1})$ and the corresponding within-group efficiency of U_2 is straightforwardly referred to as $Eff(X_{U2}^{g=2}, Y_{U2}^{g=2})$. Since the radial distance of unit U_1 to its group frontier is larger than the corresponding distance of unit U_2 compared to its group frontier (see Figure 3.4), it can be concluded that $Eff(X_{U1}^{g=1}, Y_{U1}^{g=1}) < Eff(X_{U2}^{g=2}, Y_{U2}^{g=2})$. This means that U_2 better exploits its available group technology compared to its respective counterpart from group 1.

Recall that one cannot conclude from the results presented above whether unit U_1 shows a better overall performance compared to U_2. This is because each unit is evaluated regarding its individual production frontier and, therefore, the obtained efficiency scores are not comparable as they do not refer to a common reference technology. When the researcher wants to evaluate the efficiency of each unit regarding the common metafrontier (i.e. the border shown by ABCHIJ), model (3.11) needs to be applied. The corresponding performance scores for DMUs U_1 and U_2 may be denoted as $Eff_C^M(X_{U2}^{g=2}, Y_{U2}^{g=2})$ and $Eff_C^M(X_{U1}^{g=1}, Y_{U1}^{g=1})$, respectively. Again, one can extract from Figure 3.4 that the distances of U_1 and U_2 to the global benchmark technology imply that $Eff_C^M(X_{U1}^{g=1}, Y_{U1}^{g=1}) < Eff_C^M(X_{U2}^{g=2}, Y_{U2}^{g=2})$. This means that unit U_2 also shows a better overall performance compared to U_1.

Since the metatechnology envelops both group technologies T^1 and T^2 (i.e., $PPS_C^M \supseteq PPS^{g=1}$ and $PPS_C^M \supseteq PPS_{g=2}$), it is automatically guaranteed that the efficiency values regarding the metatechnology never exceed the efficiency values measured in respect to the individual group technologies (i.e., $Eff_C^M(X_j^g, Y_j^g) \leq Eff(X_j^g, Y_j^g)\ \forall\ j = 1,...,n,\ g = 1,...G$). Furthermore, whenever an inequality between these two performance scores is observed, a so-called metatechnology gap ratio, denoted as $MTR^C(X_p^g, Y_p^g)$, can be computed as follows:

$$MTR^C(X_p^g, Y_p^g) = Eff_C^M(X_p^g, Y_p^g) \big/ Eff(X_p^g, Y_p^g). \tag{3.12}$$

The ratio expresses how close the respective group frontier is to the all-encompassing metafrontier measured at the input-output mix of DMU *p*. Hence, it indicates the technology gap between the currently available technology for DMUs in a respective group, relative to the best-observed technology available to the whole industry. This ratio can be estimated for each individual unit (see Battese and Rao 2002, p. 90). The greater the ratio is, the closer the respective DMU to the metatechnology (and vice versa). Needless to say that $MTR^C(X_j^g, Y_j^g) \leq 1$ since $Eff_C^M(X_j^g, Y_j^g) \leq Eff(X_j^g, Y_j^g)\ \forall\ j = 1,...,n,\ g = 1,...G$ is always satisfied (see Battese et al. 2004, p. 94).

Applying this idea to the case represented by Figure 3.4, one obtains metatechnology gap ratios for U_1 and U_2 satisfying $MTR^C(X_{U1}^{g=1}, Y_{U1}^{g=1}) > MTR^C(X_{U2}^{g=2}, Y_{U2}^{g=2})$. Therefore, it is straightforward to conclude that U_2 better exploits the technologies available to the whole set of groups compared to U_1.

According to (3.12), the meta-efficiency of a particular DMU can also be multiplicatively divided into two different subcomponents:

$$Eff_C^M(X_p^g, Y_p^g) = MTR^C(X_p^g, Y_p^g) \times Eff(X_p^g, Y_p^g). \tag{3.13}$$

This decomposition given by (3.13) is useful because it allows decision makers to better assess the potentials of different types of programs or business strategies to reduce the technological gap.

3.3.3 Non-convex metatechnology[27]

Recall that assuming a convex metatechnology means that weighted averages (i.e., convex combinations) of observed input-output pairs in the technology also belong to the technology set (see Hackman 2008). That is, the metatechnology is obtained by the "convex aggregation" of the group technologies (see e.g., Chen and Yang 2011, Oh and Lee 2010) and all observations from different groups are accepted to form the meta-benchmark technology. As a consequence, it is assumed that each DMU is potentially able to achieve the production possibility represented by the convex metatechnology, which is available to the whole industry in which the different DMUs operate.

However, in many cases the combination of distinct technologies is inconsistent with the primary setting of the problem by which the DMUs are partitioned to G distinct groups. Although observations in each group can be considered to be acceptable to form the respective group technology set, including all observations from all groups in the analysis (to estimate the metatechnology) is questionable. For example, the particular government rules or regulations, policy directives and economic conditions, under which the DMUs operate, can substantially differ between the distinct groups. Therefore, convex combinations of units belonging to different groups may be unrealistic and, hence, reduce the accuracy of the estimated metatechnology.

As a result, many researchers advocate the use of non-convex metafrontiers (e.g., Breustedt et al. 2008, Tiedemann et al. 2011, Sala-Garrido et al. 2011, Medal-Bartual et

[27] Excerpts of this section have been published as Afsharian, M., H. Ahn, S. G. Harms. 2018a. A non-convex meta-frontier Malmquist index for measuring productivity over time. *IMA Journal of Management Mathematics*. Vol. 29(4), pp. 377-392.

al. 2012, Huang et al. 2013 and Kerstens et al. 2019) which are formed by the pure union of all local technologies. Mathematically, this can be expressed as follows:

$$PPS_{NC}^{M} = \bigcup_{g=1}^{G} PPS^{g}. \tag{3.14}$$

The subscript "NC" in (3.14) shall highlight that the metatechnology is formed based on the non-convex union of the group technologies. Taking into account the same axioms as for the construction of the convex *PPS*, the definition of the metatechnology in (3.14) can mathematically be enhanced as follows:

$$PPS_{NC}^{M} = \bigcup_{g=1}^{G} \left\{ \begin{array}{l} (X^{g}, Y^{g}) \in \Re_{+}^{m} \times \Re_{+}^{s} \mid \; x_{i}^{g} \geq \sum_{j=1}^{\delta_g} \lambda_{j}^{g} x_{ij}^{g}, \; y_{r}^{g} \leq \sum_{j=1}^{\delta_g} \lambda_{j}^{g} y_{rj}^{g}, \\ \lambda_{j}^{g} \geq 0, \quad j = 1, \ldots, \delta_{g} \end{array} \right\}. \tag{3.15}$$

On this basis, $Eff_{NC}^{M}(X_{p}^{g}, Y_{p}^{g})$, which captures the input-oriented efficiency of unit *p* belonging to a group *g*, can be measured against any group technology *q* with the following linear programming problems:

$$Eff^{q}(X_{p}^{g}, Y_{p}^{g}) = min \left\{ \theta_{pg}^{q} \left| \begin{array}{ll} \sum_{j=1}^{\delta_g} \lambda_{j}^{g} x_{ij}^{g} \leq x_{ip}^{q} \theta_{pg}^{q}, & i = 1, \ldots, m \\ \sum_{j=1}^{\delta_g} \lambda_{j}^{g} y_{rj}^{g} \geq y_{rp}^{q}, & r = 1, \ldots, s \\ \lambda_{j}^{g,t} \geq 0, & \end{array} \right. \right\}. \tag{3.16}$$

Now with respect to the discrete nature of the metatechnology in (3.15), the meta-efficiency of DMU *p* can be computed by the following enumeration procedure:

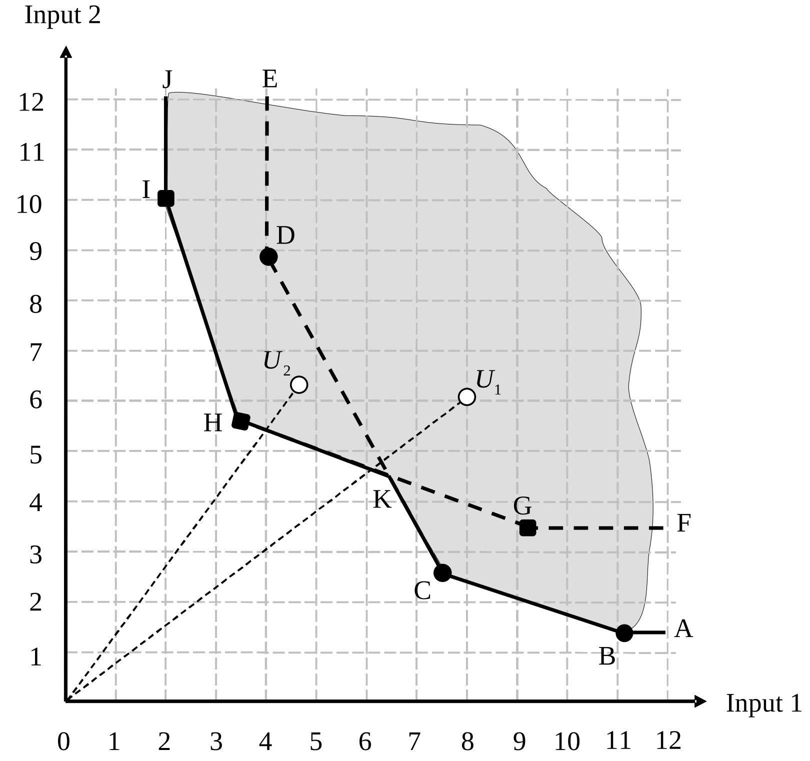
$$Eff_{NC}^{M}(X_{p}^{g}, Y_{p}^{g}) = \min_{q=1,\ldots,G} \left\{ Eff^{q}(X_{p}^{g}, Y_{p}^{g}) \right\}. \tag{3.17}$$

In this procedure, determining $Eff_{NC}^{M}(X_{p}^{g}, Y_{p}^{g})$ against the metatechnology is identical with finding the minimum value among $Eff^{q}(X_{p}^{g}, Y_{p}^{g})$ for all *q* (*q* = 1,…,*G*) in which

$Eff^{q}(X_p^g, Y_p^g)$ can also be computed in advance by the corresponding DEA models in (3.5). It should be noted that as the DMU under evaluation is a real unit, at least one of its within-group efficiencies $Eff^{q}(X_p^g, Y_p^g)$ is feasible, e.g., it is enveloped by the technology in which it has been observed. According to (3.14), this guarantees that $Eff_{NC}^{M}(X_p^g, Y_p^g)$ is feasible. However, formula (3.17) proposed above enumerates in its procedure all within-group efficiencies including those which might be unfeasible for this unit. This can occur when DMU *p* is not enveloped by the boundary of a particular group technology. For overcoming this problem in the computation of (3.17), such unfeasible results of efficiency can be replaced in advance by sufficiently big values.

To receive a better understanding of the implications of the non-convex metatechnology, revisit the computation of efficiencies based on a comparison of Figures 3.4 and 3.5. As it has been shown by Figure 3.4, the convex metafrontier is formed as the convex aggregation of T^1 and T^2. A comparison to the non-convex metafrontier shown in Figure 3.5 highlights how the convexification of these two group technologies can yield virtual points which are the pure result of the convexification of T^1 and T^2. The respective area is enveloped by points *C*, *H* and *K* (see Figure 3.4) and has never been observed in practice. In addition, a closer look at the non-convex metatechnology shows that it preserves the contribution of each group technology in the construction of the metatechnology (see Figure 3.5). In other words, information about local group technologies are not arbitrarily mixed.

Figure 3.5: Non-convex estimation of a contemporaneous technology set



According to the diagram above and the way the metatechnology is formed by the convex and non-convex approaches, it can be concluded that $PPS_{C}^{M} \supseteq PPS_{NC}^{M}$. On this basis, $Eff_{C}^{M}(X_{p}^{g},Y_{p}^{g}) \leq Eff_{NC}^{M}(X_{p}^{g},Y_{p}^{g})$ where $Eff_{C}^{M}(X_{p}^{g},Y_{p}^{g})$ and $Eff_{NC}^{M}(X_{p}^{g},Y_{p}^{g})$ denote the meta-efficiency of the convex and the non-convex approaches for a DMU *p*, respectively.

Note that – as an extreme case from a theoretical point of view – if the individual group technologies form a convex shape even in the non-convex approach, the efficiency results of these approaches will be exactly the same. However, in practical situations, the results tend to diverge because the metatechnology likely exhibits areas violating convexity in its shape.

One can form metatechnology gap ratios also for the non-convex metafrontier approach. Suppose $MTR^{NC}(X_{p}^{g},Y_{p}^{g})$ denotes the respective metatechnology gap ratio for a DMU *p* in the non-convex case. Then, the distance of the respective group frontier to the all-encompassing metafrontier can be measured at the input-output mix of DMU *p* as follows:

$$MTR^{NC}(X_{p}^{g},Y_{p}^{g}) = Eff_{NC}^{M}(X_{p}^{g},Y_{p}^{g}) \big/ Eff(X_{p}^{g},Y_{p}^{g}). \tag{3.18}$$

Again, the ratio indicates the technology gap between the currently available technology for DMUs in a respective group, relative to the best-observed technology available to the whole industry. The greater the ratio is, the closer the respective DMU to the non-convex metatechnology (and vice versa). Furthermore, one can mathematically manipulate (3.18) and decompose the meta-efficiency into two different subcomponents: The metatechnology gap ratio (which represents the respective technology gap regarding the metafrontier) and the within-group efficiency (which represents the efficiency gap regarding the local production frontier). The respective decomposition is given by formula (3.19) below:

$$Eff_{NC}^{M}(X_p^g, Y_p^g) = MTR^{NC}(X_p^g, Y_p^g) \times Eff(X_p^g, Y_p^g). \tag{3.19}$$

3.4 Malmquist index-based performance measurement

3.4.1 Background

The aforementioned approaches are useful for comparing the performance of DMUs which operate in the same time period, i.e. in a static setting. However, performance comparisons across different periods may yield further essential information for directing management decisions. For instance, one can possibly identify whether there is a productivity improvement or deterioration in the sample and indicate potential sources for performance changes. In line with these results, a DMU's management may likely improve its current performance level when it is benchmarked against other units in the data set and, subsequently, test if effective strategies can be adopted to its own operating entity.

The DEA literature comprises different techniques for dynamic performance evaluations: for example, Charnes et al. (1984) proposed the so-called Window Analysis. In this approach ,observations from different time periods are pooled and uniformly evaluated with basic DEA models. Some years later, Chambers et al. (1996) introduced the Luenberger indicator which applies directional distance functions to specify in what direction the units may be evaluated over time. However, the most-recognized approach to evaluating the productivity change of DMUs from one period to another is the so-called Malmquist index (see Malmquist 1953).

Caves et al. (1982) were the first who suggested that the Malmquist index – proposed initially for consumption analysis – can be employed in the context of performance measurement. It then took 10 years until Färe et al. (1992a) adapted the work of Caves et al. (1982) in order to DEA models for measuring productivity changes over time. In the same paper, Färe et al. (1992a) also showed how the DEA-based Malmquist index can be exhibited as the product of the technical change and efficiency change components as two important drivers of productivity changes. To date, the Malmquist index has been successfully applied to measure performance changes in different economic contexts such as banks (e.g., Berg et al. 1992), hospitals (e.g., Burgess and Wilson 1995), countries (e.g., Coelli and Rao 2005) and farms (e.g., Vassdal and Sørensen Holst 2011).

For applications in which comparing the performance of entire DMU groups is the fundamental goal, Camanho and Dyson (2006) have developed an index whose structure is also built upon the Malmquist index of Färe et al. (1992a). This performance index, however, does not measure the productivity change over a number of time periods but provides a cross-sectional comparison of the performance of groups of DMUs in a static setting. Just like the conventional Malmquist index, the performance index of Camanho and Dyson (2006) can be decomposed into various components which allows the identification of potential sources for performance differences.

According to the two aforementioned major application areas of the Malmquist index, the following subsections are structured as follows: in Section 3.4.2, the traditional Malmquist index as proposed by Färe et al. (1992a) is explained in greater detail. In the subsequent Section 3.4.3, a thorough discussion of the approach published by Camanho and Dyson (2006) is given.

3.4.2 Dynamic performance measurement[28]

Suppose that there exists a panel of *n* DMUs which have been observed in *T* (*T*>1) distinct time periods. Let the inputs and outputs of the DMUs $(X_j^t, Y_j^t) \in \Re_+^m \in \Re_+^s$ (j = 1,…,n) be denoted as non-negative and non-zero vectors $X_j^t = (x_{1j}^t, x_{2j}^t, ..., x_{mj}^t)$ and $Y_j^t = (y_{1j}^t, y_{2j}^t, ..., y_{sj}^t)$, respectively. It is also assumed that all DMUs in each period *t* operate under the same technology, resulting from, e.g., the same resource, regulatory or other environmental constraints. Hence, the contemporaneous technology in time period *t* can be represented by a *PPS* of feasible input-output combinations as follows:

$$PPS^t = \left\{ (X^t, Y^t) \in \Re_+^m \times \Re_+^s \,\middle|\, X^t \ can \ produce \ Y^t \right\}. \tag{3.20}$$

Under consideration of the same axioms as in Sections 3.2 and 3.3 (i.e., non-emptiness, free disposability and minimum extrapolation), the corresponding technology set can be specified as follows:

$$PPS^t = \left\{ \begin{array}{l} (X^t, Y^t) \in \Re_+^m \times \Re_+^s \,\middle|\, x_i^t \geq \sum_{j=1}^{n} \lambda_j x_{ij}^t, \ y_r^t \leq \sum_{j=1}^{t} \lambda_j y_{rj}^t, \\ \\ \lambda_j \geq 0, \quad j = 1, ..., n \end{array} \right\}. \tag{3.21}$$

In reference to this technology, the technical efficiency of a DMU *p* observed at time period *t* can be computed using the following programming problem:

[28] Excerpts of this section have been published in Afsharian, M., H. Ahn, S. G. Harms. 2018a. A non-convex meta-frontier Malmquist index for measuring productivity over time. *IMA Journal of Management Mathematics*. Vol. 29(4), pp. 377-392.

$$
\begin{aligned}
& Eff^{\,t}(X_p^t, Y_p^t) = min\ \theta_p^t \\
& s.t. \qquad \sum_{j=1}^{n} \lambda_j x_{ij}^t \le \theta_p^t x_{ip}^t, \qquad \forall i \\
& \qquad\qquad \sum_{j=1}^{n} \lambda_j y_{rj}^t \ge y_{rp}^t, \qquad \forall r \\
& \qquad\qquad \lambda_j \ge 0,\ \ \theta_p\ \ free\ in\ sign
\end{aligned}
\tag{3.22}
$$

whereby $Eff^{\,t}(X_p^t, Y_p^t)$ denotes the input-oriented efficiency of DMU p which has been observed at time period t. Based on these preliminaries, the Malmquist index for DMU p between two adjacent time periods t and t+1 can be formulated according to Caves et al. (1982) as

$$
MI(X_p^{t+1}, Y_p^{t+1}, X_p^t, Y_p^t) = \frac{Eff^{\,t}(X_p^{t+1}, Y_p^{t+1})}{Eff^{\,t}(X_p^t, Y_p^t)} \tag{3.23}
$$

where $Eff^{\,t}(X_p^t, Y_p^t)$ and $Eff^{\,t+1}(X_p^{t+1}, Y_p^{t+1})$ represent the efficiencies of DMU p measured in respect to the production frontiers of technology t and t+1 using formula (3.22).

Since the course of the production frontiers may change between two time periods, the computed index results depend on whether the technology in period t or the technology in period t+1 is chosen as the reference frontier (see Grifell-Tatjé and Lovell 1995, pp. 170-171). For cases where the decision maker has no preference for either of the two production frontiers, Färe et al. (1992a) proposed determining the geometric mean of two separate Malmquist indices. One Malmquist index evaluates DMU p towards the technology in t, and a second index regarding the production frontier in t+1 (see Färe et al. 1992a, p. 90). The Malmquist index of Färe et al. (1992a) thus derived can be written as:

$$
MI(X_p^{t+1}, Y_p^{t+1}, X_p^t, Y_p^t) = \left[\frac{Eff^{\,t}(X_p^{t+1}, Y_p^{t+1})}{Eff^{\,t}(X_p^t, Y_p^t)} \times \frac{Eff^{\,t+1}(X_p^{t+1}, Y_p^{t+1})}{Eff^{\,t+1}(X_p^{t+1}, Y_p^{t+1})} \right]^{1/2}. \tag{3.24}
$$

Using the technology of time period t as a reference of comparison, the first ratio in formula (3.24) measures the efficiency of DMU p observed in period t compared to the efficiency of the same DMU p observed in period t+1. The greater the ratio, the higher the obtained productivity in period t+1 compared to period t. The other direction happens when the ratio is less than 1. When the ratio signals 1, then the productivity of DMU p is similar for the two periods. The second ratio inside the brackets evaluates the same but with reference to the technology of period t+1.

Through mathematical manipulation, the Malmquist index can be further decomposed into two subcomponents that indicate potential sources of performance changes over time:

$$MI(X_p^{t+1}, Y_p^{t+1}, X_p^t, Y_p^t) =$$

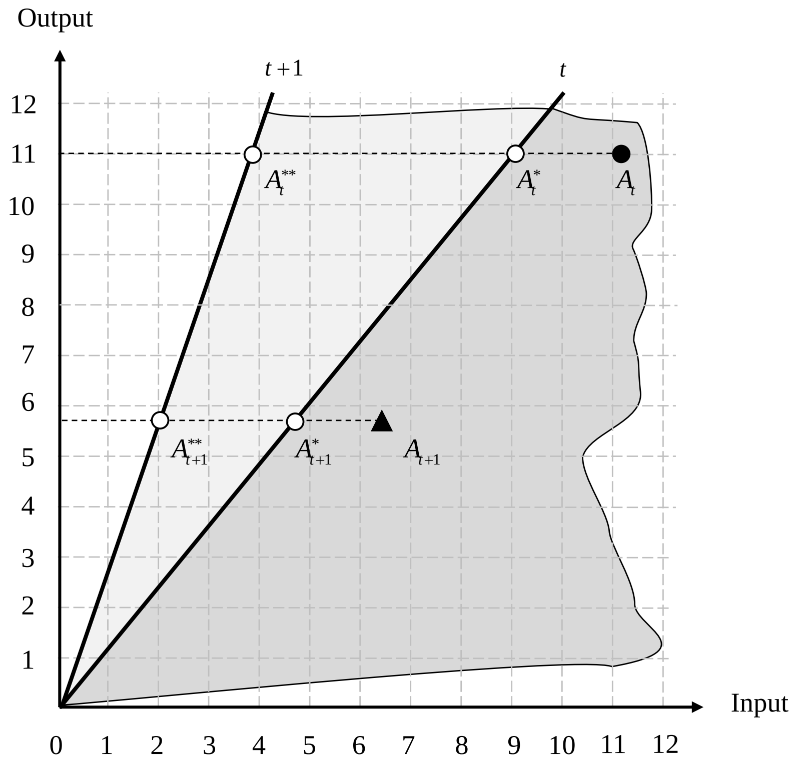
$$\underbrace{\frac{Eff^{\,t+1}(X_p^{t+1}, Y_p^{t+1})}{Eff^{\,t}(X_p^t, Y_p^t)}}_{Efficiency\ Change} \times \underbrace{\left[\frac{Eff^{\,t}(X_p^{t+1}, Y_p^{t+1})}{Eff^{\,t+1}(X_p^{t+1}, Y_p^{t+1})} \times \frac{Eff^{\,t}(X_p^t, Y_p^t)}{Eff^{\,t+1}(X_p^t, Y_p^t)}\right]^{1/2}}_{Technology\ Change}. \tag{3.25}$$

The quotient outside the bracket of equation (3.25) is called Efficiency Change (EC) and compares the distance of the DMU under assessment in each time period to that period's production frontier. That is, it measures the change of the DMU's technical efficiency score between period t and t+1. If the EC value equals 1, the DMU would obtain the same efficiency score in both periods t and t+1. A value bigger than 1 means that the DMU has become more efficient in period t+1 compared to period t. In other words, a catch-up of the DMU in respect to the period-specific production frontier has occurred. Finally, if the EC value is lower than 1, the DMU has moved farther away from the efficient boundary which corresponds to a reduction in terms of the DMU's efficiency value.

The quotient inside the bracket of equation (3.25) is called Technology Change (TC) and is a measure of the production frontier movement between period t and t+1 measured at the input-output mixes (X_p^t, Y_p^t) and (X_p^{t+1}, Y_p^{t+1}), respectively. Whereas a value over 1 represents a productivity gain by the industry, a value below 1 equals a productivity loss at the positions (X_p^t, Y_p^t) and (X_p^{t+1}, Y_p^{t+1}), respectively. Needless to say, a TC value of unity means that no production frontier-shifts occurred.

In order to receive a better understanding of the Malmquist index, revisit the TC component based on Figure 3.6. Let A_t^* signal the reference point of unit A_t on the frontier of the technology of period t and let A_t^{**} be the projection of the same observation A_t on the frontier of period *t*+1. Thus, the frontier-shift effect can be evaluated by computing the ratio between the efficiency of unit A_t in respect to technology *t* (i.e., the distance A_t - A_t^*) and the efficiency of the same observation regarding technology *t*+1 (i.e., the distance A_t - A_t^{**}). Analogously, the frontier-shift between A_{t+1}^* and A_{t+1}^{**} can be measured as the ratio of the efficiency values of observation A_{t+1} regarding the technologies *t* and *t*+1, respectively. Using the geometric mean of these two ratios, one can evaluate the frontier-shift according to Färe et al. (1992a) and, hence, the technological progress or regress represented by the respective frontier changes.

Figure 3.6: Frontier-shift of a CRS-technology



Since the traditional definition as proposed by Färe et al. (1992b) measures performance changes of DMUs by computing the geometric average of two distinct Malmquist indices, some authors criticize that it does not yield a single measure of productivity change (see Pastor and Lovell 2005). Furthermore, the traditional Malmquist index also does not satisfy the circularity property, which ensures that the index of period 1 relative to period 3 is equal to the product of the index of period 1 relative to period 2 and the index of period 2 relative to period 3 (i.e., $MI_{1,2} \times MI_{2,3} = MI_{1,3}$). If more than two periods need to be compared, this is an important requirement especially for the practical applicability of the index.

Note that the Malmquist index implicitly assumes CRS regardless of the actually present returns to scale (see Thanassoulis 2001, p. 177). Therefore, its value does not change if an output-orientation is chosen instead of an input-orientation (see Thanassoulis 2001, pp. 178-181). However, the implicit CRS assumption reduces its applicability to situations where no scale shift transformations have been applied to the data set (see Section 3.2.3). In addition, the traditional Malmquist index uses mixed-period distance functions which may cause infeasibilities when it is decomposed under VRS (see e.g., Ray and Desli 1997, Färe et al. 1997b). In response to the aforementioned drawbacks, the DEA literature comprises a huge variety of different extensions of the traditional Malmquist index (see e.g., Asmild et al. 2004, Pastor and Lovell 2005, Oh 2010, Oh and Lee 2010). A thorough overview of these approaches can be found in Afsharian and Ahn (2015).

3.4.3 Group performance comparison[29]

Based on Camanho and Dyson (2006), the performance of two groups of DMUs (e.g. group $g = 1$ and $g = 2$) can be compared by the following measure which is built upon the Malmquist index of Färe et al. (1992a):

$$PI_{2,1} = \left[\frac{\left(\prod_{j=1}^{\delta_{g=2}} Eff^{\,g=2}(X_j^{g=2}, Y_j^{g=2}) \right)^{1/\delta_{g=2}}}{\left(\prod_{j=1}^{\delta_{g=1}} Eff^{\,g=2}(X_j^{g=1}, Y_j^{g=1}) \right)^{1/\delta_{g=1}}} \times \frac{\left(\prod_{j=1}^{\delta_{g=2}} Eff^{\,g=1}(X_j^{g=2}, Y_j^{g=2}) \right)^{1/\delta_{g=2}}}{\left(\prod_{j=1}^{\delta_{g=1}} Eff^{\,g=1}(X_j^{g=1}, Y_j^{g=1}) \right)^{1/\delta_{g=1}}} \right]^{1/2} . \tag{3.26}$$

Using group technology 2 as a reference of comparison, the first ratio in formula (3.26) measures the average performance of units in group 2 compared to that of units in group 1. The greater the ratio, the higher the performance of group 2 compared to group 1. The

[29] Excerpts of this section are included in Afsharian, M., H. Ahn, S. G. Harms. 2019b. Performance comparison of management groups under centralised management. *European Journal of Operational Research.* Vol. 278(3), pp. 845-854.

other direction happens when the ratio is less than 1. When this ratio signals 1, then, on average, the performance of these two groups are similar. The second ratio inside the brackets evaluates the same but with reference to group technology 1. Following the structure of the Malmquist index, the square root of these two measures is applied in (3.26).

The use of the geometric average for the aggregation of the two ratios allows a multiplicative decomposition (see Färe et al. 1992a) into two subcomponents which may provide a better understanding of performance differences:

$$PI_{2,1} = \underbrace{\frac{\left(\prod_{j=1}^{\delta_{g=2}} Eff^{\,g=2}(X_j^{g=2}, Y_j^{g=2})\right)^{1/\delta_{g=2}}}{\left(\prod_{j=1}^{\delta_{g=1}} Eff^{\,g=1}(X_j^{g=1}, Y_j^{g=1})\right)^{1/\delta_{g=1}}}}_{Efficiency\ \ Index} \times \underbrace{\left[\frac{\left(\prod_{j=1}^{\delta_{g=2}} Eff^{\,g=1}(X_j^{g=2}, Y_j^{g=2})\right)^{1/\delta_{g=2}}}{\left(\prod_{j=1}^{\delta_{g=2}} Eff^{\,g=2}(X_j^{g=2}, Y_j^{g=2})\right)^{1/\delta_{g=2}}} \times \frac{\left(\prod_{j=1}^{\delta_{g=1}} Eff^{\,g=1}(X_j^{g=1}, Y_j^{g=1})\right)^{1/\delta_{g=1}}}{\left(\prod_{j=1}^{\delta_{g=1}} Eff^{\,g=2}(X_j^{g=1}, Y_j^{g=1})\right)^{1/\delta_{g=1}}}\right]^{1/2}}_{Frontier\ \ Index} . \tag{3.27}$$

The Efficiency Index (EI) compares the within-group efficiency spreads. Therefore, it relates the average efficiencies of DMUs from group 1 and 2 to their respective technologies. A value less than one indicates that the efficiency spreads of DMUs in group 2 are smaller than the efficiency spreads of DMUs in group 1. From its mathematical composition, it follows that the EI does not allow any interpretation of productivity advantages in either group. This is because the index uses different reference technologies for the comparison in the numerator and denominator, respectively. The Frontier Index (FI) inside the square root bracket uses the frontier of group 1 as a reference for DMUs of group 1 and 2 in the numerator. In the denominator, the technology of group 2 is used to evaluate the DMUs of both groups. Thus, the FI compares the distances between the respective boundaries of group 1 and 2. An FI larger than unity indicates that DMUs forming the boundary in group 2 are (on average) more productive than their respective counterparts in group 1. A value below unity proves that the DMUs forming the boundary of group 1 are (on average) more productive than the frontier DMUs of group 2.

By using individual reference frontiers for each DMU, the index of Camanho and Dyson (2006) is limited to environments with CRS only. If VRS is assumed, some DMUs may not be able to be projected onto the respective reference frontier. In these cases, the approach of Camanho and Dyson (2006) yields infeasible results. Due to multiple reference technologies, the approach of Camanho and Dyson (2006) does not fulfil circularity either (i.e. $PI_{2,1} \times PI_{3,2} = PI_{3,1}$) which would guarantee that a ranking of groups according to their corresponding performance values is possible. Since circularity is not fulfilled and comparisons of individual performance values are not possible, the index may end up with counter-intuitive results that hamper managerial interpretations.

In order to cater to circularity, Camanho and Dyson (2006) proposed an extension of their index. The adjusted index is based on the idea that the circularity problem is caused by different reference technologies needed for the calculation of the FI. Therefore, Camanho and Dyson (2006) proposed a new FI which involves the distance between any two frontiers at the input-output mixes of all groups under comparison. For the case that G distinct groups are involved in the evaluation, the new performance index between groups 1 and 2 is as follows:

$$PI_{2,1}^{Adj} = \underbrace{\frac{\left(\prod_{j=1}^{\delta_{g=2}} Eff^{\,g=2}(X_j^{g=2}, Y_j^{g=2})\right)^{1/\delta_{g=2}}}{\left(\prod_{j=1}^{\delta_{g=1}} Eff^{\,g=1}(X_j^{g=1}, Y_j^{g=1})\right)^{1/\delta_{g=1}}}}_{Efficiency\ \ Index} \times \underbrace{\left[\prod_{q=1}^{G} \frac{\left(\prod_{j=1}^{\delta_q} Eff^{\,g=2}(X_j^{q}, Y_j^{q})\right)^{1/\delta_q}}{\left(\prod_{j=1}^{\delta_q} Eff^{\,g=1}(X_j^{q}, Y_j^{q})\right)^{1/\delta_q}}\right]^{1/G}}_{Adjusted\ \ Frontier\ \ Index} . \tag{3.28}$$

As equation (3.28) shows, the EI of the new performance index is identical with the index proposed in formula (3.27), because this component already satisfies circularity. The adjusted frontier index (AFI) in the brackets of equation (3.28) incorporates all DMUs under evaluation into the frontier comparison of groups 1 and 2 and, thus, guarantees the circularity of this component. Therefore, the index provides a robust performance ranking of groups of DMUs operating under different technologies (see Camanho and Dyson 2006, p. 41).

4 Organizational structures and their modeling in DEA

4.1 Introduction

In order to receive meaningful performance scores, DEA models need to be adapted according to their respective application area. For example, the selected input and output factors have to correctly represent the DMUs' transformation processes (see Golany and Roll 1989).[30] It is also necessary to estimate the PPS. Therefore, researchers have to make assumptions about convexity (see Kerstens et al. 2019)[31] and returns to scale which characterize the production technology (see Banker et al. 2004).

The aforementioned assumptions have in common that they are substantially based on the perspective of traditional production theory as the foundation of DEA (see e.g., Dyckhoff 2006, p. 2, Dyckhoff and Ahn 2010, p. 1252). However, the discussions in Chapter 1 show that DEA models should also account for different organizational structures. For example, basic DEA models (e.g., the CCR or BCC models) may project inefficient DMUs onto the efficient frontier in such a way that the resulting targets may be unrealistic if the top management does not allow a specialization on certain input or output factors. In such cases, the DMUs are required to pursue more "balanced" targets.[32] One may also think of situations where the performance of subunits needs to be evaluated according to

---

30 See for example the approach of Ahn and Le (2014) who apply a goal-oriented framework to systematically derive input and output factors for the case of German savings banks.

31 See also the discussions about convex and non-convex metatechnologies in Sections 3.3.2 and 3.3.3, respectively.

32 This topic has been addressed by the approaches proposed by Ahn et al. (2012), Dyckhoff et al. (2013), Dyckhoff and Gutgesell (2015) as well as Ahn and Vazquez Novoa (2018).

an overall business strategy. This requires that target values are derived according to managerial specifications (see. e.g., Chambers et al. 1996).[33]

The given examples indicate how different organizational settings influence the way performance has to be measured. Against this backdrop, it is somehow surprising that only little effort has been undertaken to align ideas of organization theory and DEA.[34] In consideration of this research gap, this chapter has a twofold aspiration: on the one hand, a brief overview of so-called organizational variables is given. These variables are used in the respective theory to describe an organization's structure. On the other hand, a systematic literature review shows to what extent the concept of centralization – one of the most frequently discussed organizational variables – is already modeled in current DEA approaches. The limitation to this organizational variable is necessary to receive a manageable amount of relevant research papers. However, the selection is also reasonable because of its special relevance for the case of KONE Corporation.

The rest of the chapter unfolds as follows: in Section 4.2, five important organizational variables are described. The concept of centralization is explained in more detail from the perspective of DEA in the subsequent Section 4.3. Special emphasis is placed on deriving appropriate definitions that allow classifying the DEA approaches according to the way in which different degrees of centralization are modeled. Based on the respective classification, a systematic literature review is provided in Section 4.4. The chapter concludes with Section 4.5, which elaborates detailed research questions to be answered in the remaining course of this thesis.

---

[33] A predetermined direction vector of the so-called Luenberger indicator can straightforwardly model this idea. See Afsharian and Ahn (2014) for a comprehensive overview of this approach.

[34] For example, the overview of Mar-Molinero et al. (2014) shows that relatively little research has been conducted on modeling different degrees of centralization. In fact, the authors only identify eight such approaches (see Mar-Molinero et al. 2014, p. 275).

4.2 The perspective of organization theory

A series of empirical studies have proven a significant connection between a company's performance and the underlying organizational structure.[35] This has facilitated the scientific consensus that a company's performance strongly depends on the design of certain organizational variables. However, widespread disagreement exists regarding their suitable definition and categorization. For example, a literature review of Cordes-Berszinn (2013) shows that there are at least 19 different classification schemes of organizational variables. As the intention of this thesis is not to provide a solution to this scientific dis agreement, the following explanations concentrate on the five most frequently used variables, which are according to Cordes-Berszinn (2013): specialization, coordination, centralization, configuration and formalization.[36]

4.2.1 Specialization

In complex organizations, an individual organization member is no longer able to oversee and conduct all tasks efficiently in a cost- and quality-oriented way. To solve this problem, it is necessary to divide a superior task into several subtasks and allocate them to distinct organization members (see Kogelheide 1992, p. 246). This division of labor is usually referred to as "specialization" (see Kieser and Walgenbach 2010, p. 73).

The subtasks can be allocated according to different criteria, like objects (e.g., regions, customers, products) or functions (e.g., procurement, production, selling) (see Weinert 2002, p. 13). A major advantage of specialization is that the distribution of subtasks can be adapted to the respective knowledge and abilities of each organization member. A high degree of specialization may also lead to learning effects or economies of scale. In the end, these advantages may raise efficiency and effectiveness gains for the entire organization. However, extreme subdivision of tasks may also lead to an excessively narrowed perspective for each individual. Furthermore, too specialized organization members may be unsatisfied with their task diversity, which usually causes indifference, demotivation and low-quality work results (see Bea and Göbel 2010, pp. 290-291).

---

35 See Hao et al. (2012) for a review of the relationship between organizations' structures and performance.

36 The mentioned variables are identical with the ones used by Kieser and Walgenbach (2010).

4.2.2 Coordination

When an organization makes use of the concept of specialization, this may correspondingly cause a high degree of work interdependencies. For example, an organization member working at a later stage of the production process may substantially rely on the respective work quality at a previous stage of the production process (see Kieser and Walgenbach 2010, p. 93). Such relationships typically require coordinating the individuals' behavior in line with the overall organizational objectives (see Ewert and Wagenhofer 2014, p. 387). This so-called coordination can be achieved based on different instruments such as profit centers or transfer prices (see Bea and Göbel 2010, p. 297).

If coordination mechanisms are not sufficiently implemented, the building of realistic expectations regarding the behavior of organization members is affected. In the best case, coordination may allow all organizational activities to be controlled and predicted with high accuracy. Thus, the possibility that expectations regarding the behavior of individual members meet the reality is greater (see Cordes-Berszinn 2013, p. 125). However, increased coordination usually causes numerous communications among the involved parties and, thus, bureaucratic processes. Furthermore, coordination instruments are not free of cost (see Cray 1984, p. 87, Demski 1997, pp. 583-587) and must be reasonably aligned (e.g., when performance measures are used in combination with incentive systems). Otherwise, coordination problems between different parties may occur which compromise the achievement of the overall organizational objectives. Therefore, one can expect that partial coordination is the usual case (see Demski 1997, p. 587).

4.2.3 Centralization

According to Mintzberg (1979), the term "centralization" has been so frequently applied that it has "almost ceased to have a useful meaning''. For example, Graubner (2006) mentions that the geographical distribution of an organization has been called decentralization in the non-English literature.[37] In contrast to that, the term centralization is defined in the context of this thesis by the respective locus of authority to make decisions affecting the

[37] In line with this, Faust et al. (1994) also mention that the term „decentralization" is used in different ways. They use the term "operative decentralization" to describe the idea of "delegation" (see Faust et al. 1994, p. 23).

organization (see Pugh et al. 1968, p. 76, Krasman 2011, p. 16).[38] When (a certain kind of) decision authority is located on a hierarchically low level, one usually speaks of a "decentralized organization" (with respect to the kind of authority in question). By contrast, when the locus of authority corresponds to a high hierarchical level, this characterizes a "centralized organization" (see Daft et al. 2010, p. 18).

A shift towards centralization (i.e., a decrease in the degree of decentralization) may generate economies of scale from resource allocations between different organizational members. Some authors also claim that centralization improves the quality of planning and control (see Hammann 1976). This, however, strongly depends on the respective situation. For example, it is problematic if the centralized management does not have adequate information or the required competence for making effective decisions.

Correspondingly, a shift towards decentralization can have a positive effect on the quality of decisions. Furthermore, the increased decision authority on the lower hierarchical levels can also improve the motivation of the individual organizational members since not all decisions require previous approval by the top management. In the end, this can also enhance the flexibility of the overall organization (see Picot 1993, p. 222). However, if an organization wants to make use of decentralization, it is necessary to carefully select which decision authority can be located on lower hierarchy levels and which should remain at the hierarchical top-level (see Koontz and Weihrich 2010, pp. 183-184). In fact, a complete decentralization of all decision making authorities implies that there might be no overall organization at all because every entity is operating according to its individual preferences (see Koontz and Weihrich 2010, pp. 183-184). Additional costs caused by non-transparent processes, duplication of work or extensive coordination may be a result of this decision autonomy (see Picot 1993, Weber and Gschmack 2012). Therefore, one usually observes a compromise between the dichotomous forms of a clearly decentralized and a clearly centralized organization in practice.[39]

---

[38] Similar definitions are also given by Fredrickson (1986, p. 282), Burton and Obel (2004, p. 80), Pleshko (2007, p. 54) as well as Willem and Buelens (2009, p. 152).

[39] In Section 4.3.3, the term "hybrid management" is introduced for these organization types. The word "hybrid" emphasizes that the respective organization can be neither classified as clearly centralized nor decentralized.

4.2.4 Configuration

Kieser and Walgenbach (2010) refer to the three aforementioned variables as "central principles or mechanisms". However, the authors emphasize that a consistent description of each organization should also include the exterior shape of the role structures. These role structures, called "configuration", are usually depicted by organization diagrams (see Pugh et al. 1968).[40]

The degree of configuration is substantially influenced by the competences of an organization's members to decide and to direct other members (see Kieser and Walgenbach 2010). Therefore, Kieser and Walgenbach (2010) refer to configuration also as a "system of guidance and control". In practice, one may meet three distinct types of configurations: functional, divisional and matrix organizations which differ regarding the degree and kind of specialization on the second hierarchy level as well as the degree of centralization and the form of coordination (see Bea and Göbel 2010, p. 311). The different types of configurations determine "who" is instructing "whom" and "who" gets instructions from "whom". Correspondingly, a totally configured organization means that all information channels are fully dictated, which can substantially reduce the flexibility of the organization. In contrast, if there were no configurations and, hence, clear definitions of "who" is instructing "whom", a chaotic and inefficient organization would be the logical result (see Cordes-Berszinn 2013, p. 125).

4.2.5 Formalization

The term "formalization" refers to the "extent to which an organization uses rules and procedures to prescribe behavior" (see Fredrickson 1986, p. 283).[41] In other words, formalization explicitly defines for each member how a given task has to be conducted. This also includes "where" and "from whom" the task is performed (see Fredrickson 1986).

---

[40] The term "configuration" is a crucial subject of discussion in literature. As this is not the focus here, the interested reader is referred to the overview of different definitions of the term "configuration" given by Schulte-Zurhausen (2014, p. 247).

[41] In many cases, the definition of "formalization" is somewhat extended. For example, Pugh et al. (1968) mention that corresponding specifications need to be present in written form. However, it is also argued that such extensions of the definition mentioned here are usually applied to simplify the measurement of formalization in empirical studies (see Fredrickson 1986).

The rules and procedures can be numerous and finely tuned or be few and less aligned. Organizations, which are characterized by a high degree of formalization, have the advantage that the leadership can obtain control over the organization. This can cause efficiency gains since redundancy of tasks or responsibilities can be largely eliminated through standardized behavior. However, it also restricts individual autonomy. Consequently, the (dis-)advantage of formalization is substantially depending on the respective organization department. For example, a production department producing standardized articles is typically characterized by a high degree of formalization. In contrast, a R&D department which requires flexibility and creativity usually has a small degree of formalization (see Burton and Obel 2004, p. 78).

4.3 The perspective of DEA

The aforementioned variables constitute central terminologies of the organization theory. Some of them have also been applied in the field of DEA. A brief overview of how the concept of centralization is specified in the traditional DEA literature is provided in the following subsections. Concentrating on this organizational variable is appropriate to provide a profound theoretical basis for the literature review in Section 4.4. In addition, the concept of centralization has been modeled in a variety of recent DEA publications (e.g., Lozano 2014, Mar-Molinero et al. 2014) without being adequately defined. The discussions in Section 2.1 have also shown that substantial characteristics of a centralized management can be found in the case of KONE’s regional management concept, which is used to coordinate the different maintenance units. Hence, this aspect should be sufficiently considered in the yet to be proposed performance measurement framework.

4.3.1 Decentralized management

As it has been introduced in Section 3.1, the units under evaluation are typically referred to as “decision making units” in the context of DEA. This term has already been used in the seminal work of Charnes et al. (1978) and implies that each DMU is characterized by an extensive degree of autonomy in making decisions. The characterization of the unit under assessment as "decision making" also means that it has control over its operating process that transforms its inputs into a bunch of outputs (see Thanassoulis 2001, pp. 21-

22). Besides, the DMUs are assumed to operate independently from each other and collaborations between distinct units are entirely neglected. In consequence, the DMUs can be considered as competitors rather than companions (see Lozano and Villa 2004, pp. 143-144).

The underlying assumption of basic DEA models that each DMU has the *authority to make decisions* according to its individual preferences shows substantial similarities to the definition of a decentralized organization as provided in Section 4.2. Recall that basic DEA models allow each DMU to choose freely an individual set of weights,[42] which puts the respective unit in the best possible light (see Cooper et al. 2011b, p. 21). In other words, these basic DEA models implicitly incorporate major characteristics of a decentralized organization by providing weight flexibility to each DMU and, therefore, allow them to make decisions (e.g., in terms of resource usage) according to their individual preferences (see Afsharian et al. 2018b).

Hence, a DEA approach based on the assumption of a decentralized management scenario is characterized by the following definition:

A DEA approach is called "decentralized" if it (i.e., the respective mathematical models) consistently allows each DMU to choose an individual set of weights.

Note that numerous DEA approaches comprise multiple steps and apply two or more programming problems in a subsequent manner. The term "consistently" in the definition above accounts for this problem and implies that a DEA approach is only called decentralized when an individual set of weights is applied *throughout each step* (see e.g., the slack-based DEA approaches in Charnes et al. 1978 and Banker et al. 1984).

4.3.2 Centralized management

The assumption of a decentralized management scenario is not always applicable, since many organizations do not make comprehensive use of decentralization. In several instances, the DMUs are required to hand over a substantial share of their decision making authority to a central decision maker who decides according to its own preferences. The

[42] See the corresponding programming problems in (3.4) and (3.6).

DEA literature typically refers to these business environments as “centralized management scenarios” (see Afsharian et al. 2018b, p. 2).

In such scenarios, one of the major DEA assumptions is that the centralized management has the ability to reorganize the allocation of resources (see e.g., Athanassopoulos 1995, Färe et al. 1997a, Beasley 2003), which is also a substantial similarity to the central decision maker as described in the literature on organization theory (see e.g., Hammann 1976). In centralized management scenarios, it is also supposed that the individual DMU has no control over resource usage and output production (see Afsharian and Ahn 2017), which is a fundamental assumption of decentralized DEA approaches.

Typically mentioned examples of centralized management scenarios in practice are stores of a pharmacy chain (see e.g., Ahn et al. 2012), hospitals operating under the umbrella of a company headquarters (see e.g., Mar-Molinero et al. 2014) and centrally coordinated fire departments (see e.g., Fang and Zhang 2008). In these situations, the central decision maker is not interested in increasing the performance of each subordinated unit. Instead, he aims at a performance improvement of the complete set of DMUs and, hence, the entire organization.[43]

To appropriately model centralized management scenarios, various DEA models have been proposed over the last few years (see e.g., Lozano et al. 2004, Lozano and Villa 2005, Kao and Hung 2005, Varmaz et al. 2013). All these approaches have in common that they explicitly restrict the ability of each DMU to choose an *individual* set of weights. Instead, a *common* set of weights is applied to evaluate the entire set of DMUs.

The economic interpretation of this mathematical modification is as follows: since the DMUs are not able to choose the weights according to their own preferences, they are projected onto the efficiency frontier in a mutual manner (see Lozano and Villa 2004, p. 144). In other words, the DMUs are aggregated within the modeling process and, hence,

[43] Note that it is not a necessary requirement that all units operate within the same organization. Decision making authorities with the abilities mentioned here also occur in inter-organizational contexts. For example, national governments have considerable influence on the decisions, strategies, objectives and resources in network industries such as electricity, natural gas, water supply and telecommunication. This is achieved on the basis of comprehensive regulation mechanisms which manage the behavior of the operators, ensuring that appropriate services at reasonable prices are delivered to the customers (see Afsharian et al. 2019a, p. 1).

considered as a corporate multi-unit organization. The common set of weights in turn represents the preference structure of the central decision maker that is enforced to the DMU level. It also means that collaborations and resource allocations are explicitly incorporated in the modeling process (see Lozano and Villa 2004, p. 144).

In order to distinguish between the decentralized DEA approaches as defined in Section 4.3.1 and the approaches applicable in centralized management scenarios, the following definition is proposed:

A DEA approach is called "centralized", if it (i.e., the respective mathematical models) consistently applies a common set of weights to all DMUs.

Similar to the definition of decentralized DEA approaches, the term "consistently" emphasizes that the idea of a common set of weights has to be straightforwardly adapted within each step of the respective approach (see e.g., Lozano and Villa 2004, Lozano and Villa 2005).

4.3.3 Hybrid management

Note that the previously mentioned DEA approaches are only applicable to quite extreme cases of the degree of organizational centralization (i.e., complete decentralization or complete centralization). In practice, there are relatively few companies that can be either classified as clearly decentralized or clearly centralized. The majority of organizations usually shows (at least to some extent) characteristics of both concepts. These cases are referred to as "hybrid management scenarios" in the remaining chapters of this thesis.

Despite its obvious practical relevance, it is interesting to note that only Afsharian et al. (2019c) *explicitly* model hybrid management scenarios using DEA.[44] However, in line with the definitions established in Sections 4.3.1 and 4.3.2, there are numerous other approaches that *implicitly* model hybrid management scenarios (see e.g., Roll et al. 1991, Roll and Golany 1993, Kao and Hung 2005): at some point of the respective procedure, these approaches allow each DMU to choose an individual set of weights, and at some

[44] To the best of our knowledge, the idea of different degrees of centralization is only mentioned by Afsharian et al. (2019c).

other point, they apply a common set of weights. In other words, these approaches use a hybrid specification of weights.

Correspondingly, a DEA approach based on the assumption of a hybrid management scenario is characterized here by the following definition:

A DEA approach is called "hybrid" if it (i.e., the respective mathematical models) is neither fully decentralized nor fully centralized.

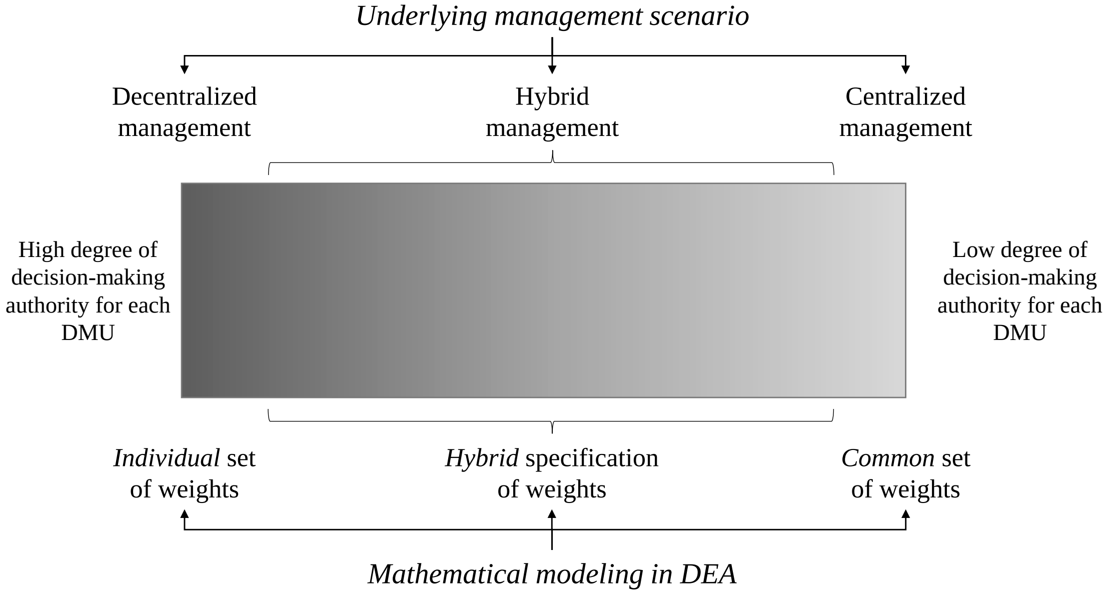
In line with the above definition, approaches are classified as "hybrid" if they neither consistently apply an individual nor a common set of weights. Accordingly, for example, two-stage approaches that use an individual set of weights in a first step and a common set of weights in a second step (or vice versa) are classified as hybrid (see e.g., Roll et al. 1991, Roll and Golany 1993).

4.4 A systematic literature review of hybrid and centralized DEA approaches

There are already several systematic literature reviews that report the current state-of-the-art of DEA approaches. The majority of reviews are quite general (see e.g., Emrouznejad et al. 2008, Emrouznejad and Yang 2018) or focus on approaches with an underlying decentralized management scenario. However, as it has been indicated in Section 4.3, there are various situations in practice where the respective organization is characterized by a hybrid or centralized management. Nevertheless, there are no publications that provide a comprehensive and systematic overview of respective approaches proposed in literature.[45] To close this gap, the following review exclusively focuses on hybrid and centralized DEA approaches. The respective definitions elaborated in Section 4.3 above are summarized in the following Figure 4.1.

[45] A brief discussion of different centralized DEA approaches is only given by Mar-Molinero et al. (2014). However, this publication does not provide a profound review.

Figure 4.1: Distinction between the three management scenarios[46]

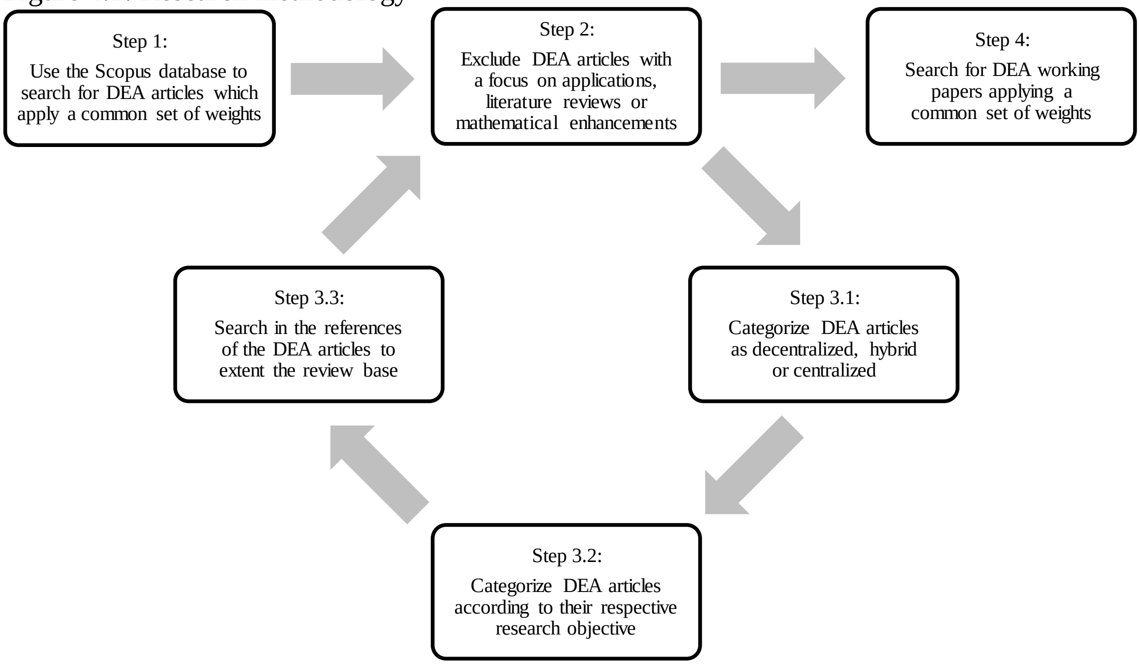


4.4.1 Research methodology

The current literature was systematically reviewed using the research methodology depicted in Figure 4.2. This methodology consists of four steps, whereby the third one includes three substeps according to the three subjects of the content analysis.

[46] This Figure is based on Horngren et al. (1996, p. 373) and has been expanded by the lower part.

Figure 4.2: Research methodology



In Step 1, a comprehensive internet search was conducted. The database Scopus was chosen for this enquiry as it is often considered as the largest database of peer-reviewed scientific literature (see Scopus 2018). In order to identify as many potentially relevant publications as possible, the search was based on the combination of four keywords: “data envelopment analysis”, “Common”, “Set” and “Weights”. Using a specific configuration of Scopus, the search engine was asked to look for the aforementioned keywords within the full document (i.e., in the title, abstract, keywords, text and bibliography of each publication). In order to reasonably confine the results, additional filter criteria were applied: First, the search with Scopus was limited to publications of the subject areas “Decision Sciences”, “Business, Management and Accounting”, “Social Sciences” and “Economics, Econometrics and Finance”. Hence, subject areas like “Chemistry”, “Engineering” or “Mathematics” etc. were excluded, since publications within these categories usually show a weak link to the field of business sciences. Second, the internet search was also restricted to “Journal articles” only, because it can be plausibly reasoned that the majority of recognized developmental DEA approaches is published in scientific journals. Consequently, monographies, book chapters as well as conference proceedings were not included in this study.[47] Third, the literature search focused on English literature only.[48] In this way, 226 potentially relevant articles were identified.

In step 2, the set of articles was gradually reduced by means of reviewing their content. After the exclusion of obviously irrelevant contributions according to their title and abstract, the remaining publications were screened with respect to their kind of contribution. For example, many articles exclusively comprise empirical applications or literature reviews of already suggested DEA models; others only provide mathematical adaptions, minor enhancements or computational simplifications. Such publications were also excluded in order to obtain only articles containing unique and novel DEA approaches.

The remaining publications were then thoroughly analyzed in Step 3. First, they were categorized according to the underlying management scenario (Substep 3.1). To this effect, the proposed mathematical models as well as their empirical applications were compared with the definitions deduced in Section 4.3. A major challenge of this step was caused by the mathematical formulations of the respective approaches: since many articles do not represent both dual DEA formulations – the multiplier *and* envelopment forms (see Section 3.2) – a simple comparison with the definitions of Section 4.3 was not always possible. In some cases, the authors only present the envelopment form. In order to also categorize these articles, the mathematical relationship was used that an objective function evaluating the entire set of DMUs in the *envelopment* form of a DEA model corresponds to a common set of weights in its *multiplier* form (Lozano and Villa 2004). Thus, it was possible to classify all approaches as decentralized, hybrid or centralized by reviewing the respective weights and/or objective functions (regardless of the mathematical problems having been formulated in the multiplier or envelopment form).

---

[47] This restriction is in line with other highly recognized literature surveys (see e.g., Emrouznejad and Yang 2018).

[48] The entire search string that has been used for the Scopus-based internet search is as follows: ALL ("Data Envelopment Analysis" AND common AND set AND weight) AND (LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "ECON")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")).

The DEA approaches classified as hybrid and centralized in Substep 3.1 were then further categorized in the course of Substep 3.2 according to the research objectives of the publications. In other words, the task was to examine the authors' intention for their application of a certain weighting scheme. This analysis provided a better understanding of the respective approach, its potential application areas and, in addition, insight whether it is appropriate for the special case of KONE. Ultimately, the identified research objectives were aggregated into eight major categories, which are called "research streams" in the following.

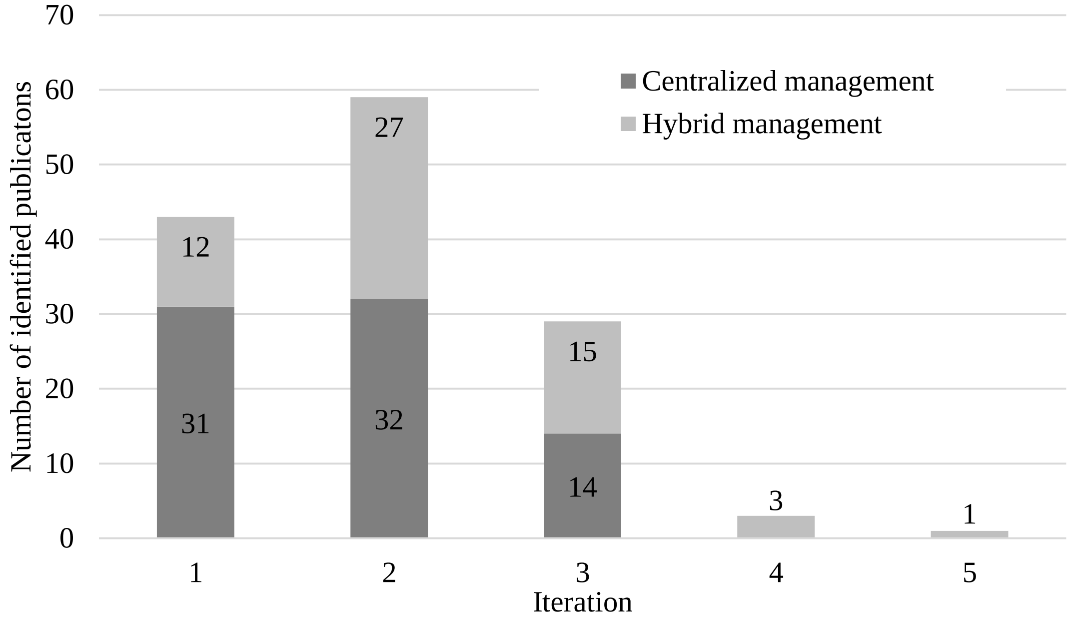
Note that the use of Scopus as the only search engine for identifying related publications (see Step 1) is in line with the research methodology of highly recognized articles to identify exhaustive DEA bibliographies (see e.g., Emrouznejad and Yang 2018). However, one may argue that using only a single database stands in clear contrast to other recently published systematic literature reviews and can raise criticism especially regarding the completeness of the bibliography.[49] Furthermore, the keywords used in Step 1 aim at identifying DEA approaches where the authors *explicitly* mention that a common set of weights is applied. However, there may also be approaches that apply a common weighting scheme *implicitly* and, consequently, also match the definition of a hybrid or centralized DEA approach as elaborated in Section 4.3.

Being aware of these inherent methodological issues, Substep 3.3 had the purpose of broadening the search for hybrid or centralized DEA approaches. To this end, the references of all publications already determined as relevant were reviewed in detail. The publications thus identified as also meriting analysis were then subject to Step 2 and 3. These steps were conducted in an iterative way until no further publications to be analyzed could be found. In the current study, this status was achieved after five iterations, leading to a number of 135 publications – 77 contributions relatable to centralized management and 58 contributions relatable to hybrid management. Figure 4.3 illustrates the respective result.

---

[49] For example, Ahn et al. (2018a) use ten different online databases (EBSCO, ECONBIZ, Emerald Insight, GVK PLUS, JSTOR, SAGE Journals, Science Direct, Springer Link, Wiley Online Library, WISO) to identify the current state-of-the-art of target costing methods. To provide a DEA bibliography, Gattoufi et al. (2004) used six databases (ABI, ECONLIT, Science Direct, JSTOR, Kluwer Verlag and Wiley Inter Science), and Emrouznejad et al. (2008) used five (Science Direct, EBSCO, Google Scholar, JSTOR and Pro-Quest).

Figure 4.3: Number of identified approaches by the iterations of Step 2 and 3



The purpose of Step 4 – the last phase of the research methodology – was to also incorporate most recent contributions into the literature review. Such newly developed approaches are typically not published in journal articles yet and, therefore, are ignored by a Scopus-based internet search. Instead, many authors publish their recent findings as working papers. Therefore, an internet search was conducted in December 2018 on the Social Science Research Network website (see SSRN 2018) as well as on selected university websites.[50] However, no relevant working papers could be identified.

4.4.2 Research streams and their characteristics

4.4.2.1 Overview

Based on Step 3.2 of the research methodology, eight distinct research streams were identified. They are listed in Column 1 of Table 4.1 according to the chronological appearance of the most influential article of the respective research stream (see Table 4.2). The num-

50 Note that the university websites of the six authors with the most publications on hybrid and centralized DEA approaches have been taken into account within this internet search. These "most productive authors" are A. Amirteimoori, W. D. Cook, L. Liang, F. H. Lotfi, M. Toloo and Y.-M. Wang (see Section 4.4.3).

ber of allocated hybrid or centralized DEA approaches are presented in Columns 2 and 3, respectively. Column 4 shows the total number of publications for each research stream, while Column 5 indicates their share in percent.

Table 4.1: Research streams and the respective number of DEA approaches

		Underlying management scenario			
		Hybrid management	Centralized management	Sum	Share in %
Research stream	1 Controlling factor weights	12	13	25	18.5
	2 Classification schemes	3	2	5	3.7
	3 Resource allocation & target setting	16	22	38	28.2
	4 Ranking of DMUs	19	21	40	29.6
	5 Improving the discrimination power	1	2	3	2.2
	6 Finding the (single) most efficient DMU	2	13	15	11.1
	7 Construction of composite indicators	2	4	6	4.4
	8 Dynamic performance measurement	3	0	3	2.2
Sum		**58**	**77**	**135**	**100**

Interestingly, the different streams partially correspond to the limitations of basic (i.e., decentralized) DEA approaches discussed in Section 3.2.4. This clearly indicates that many authors consider the application of common or hybrid weighting schemes as a potential way to avoid fundamental problems associated with traditional DEA models. For example, the research stream *Controlling factor weights* focuses on the critique that basic DEA models often assign extreme weights to certain input-output factors (see e.g., Dyson and Thanassoulis 1988). Another drawback of basic DEA models is that they may show a low discrimination power, especially when the number of observations is small in comparison to the number of selected inputs and outputs (see e.g., Angulo-Meza and Lins 2002, Adler and Yazhemsky 2010). Different approaches have also been proposed for this problem which are based on the application of a common set of weights. These publications have been classified as research stream *Improving the discrimination power*.

Table 4.1 shows that the vast majority of hybrid and centralized DEA approaches falls into the categories *Resource allocation & target setting* and *Ranking of DMUs*. Together, both research streams account for around 58 % of all publications found. By contrast, the five smallest research streams in terms of number of publications (e.g., *Improving the discrimination power*, *Construction of composite indicators*, *Finding the (single) most* *efficient DMU*, *Dynamic performance measurement* and *Classification schemes*) only comprise around 24 % of the publications.

The following subsections are structured according to the research streams and their order in Table 4.1. In each subsection, the most influential hybrid or centralized DEA paper (i.e., the paper with the most citations from other papers proposing hybrid or centralized DEA approaches) is explained. Due to the vast amount of identified publications, it is necessary to focus on these selected papers, which are listed in Table 4.2. However, note that a complete list of all papers focusing on hybrid and centralized DEA approaches is included in the Appendix of this thesis (see page 208).

Table 4.2: Overview of the most cited approaches in each research stream

		Most cited approach	Underlying management scenario
Research stream	1 Controlling factor weights	Roll et al. (1991)	Hybrid
	2 Classification schemes	Sueyoshi (1999)	Hybrid
	3 Resource allocation & target setting	Beasley (2003)	Hybrid
	4 Ranking of DMUs	Kao and Hung (2005)	Hybrid
	5 Improving the discrimination power	Karsak and Ahiska (2005)	Centralized
	6 Finding the (single) most efficient DMU	Amin (2009)*	Centralized
	7 Construction of composite indicators	Hatefi and Torabi (2010)	Centralized
	8 Dynamic performance measurement	Kao (2010)	Hybrid

* Note that Amin and Toloo (2007) is the most cited paper of research stream 6. However, due to inherent mathematical flaws in this article, a corrected version published by Amin (2009) is described in the respective subsection.

4.4.2.2 Controlling factor weights

The flexibility of basic DEA approaches allows each DMU to choose a set of weights, which presents its performance in the best possible light. In specific cases, this extensive flexibility causes extremely high or extremely low (even zero) weights for some input or output factors. However, many authors have criticized such extreme weighting schemes, especially if the performance regarding certain input or output factors should not even partially be ignored in practice. For example, when assessing different university departments with the two outputs “number of graduated master students” and “number of graduated bachelor students”, it might be hard to justify that a university department assigns zero weights to one of the two outputs when governmental regulations demand the education of both student types. As a result, the relative efficiency of a DMU may be flawed and not appropriately reflect its “real” performance. In specific cases, this means that a certain DMU may be evaluated as relatively efficient merely because its ratio for a certain, possibly irrelevant, input-output combination is the highest when compared to the remaining DMUs in the data set (see Dyson and Thanassoulis 1988, p. 564). Furthermore, it may also be difficult to argue that widely differing weights are attached to the same factor only because different units are evaluated (see Roll et al. 1991, p. 3).

In order to overcome this problem, some authors have proposed approaches where constraints on weights are used (see e.g., Podinovski and Athanassopoulos 1998, p. 564, Podinovski 2004a, Podinovski 2004b). These so-called weight restrictions (or weight bounds) reduce the ability of DMUs to choose extreme weights. However, some authors have also suggested that a common weighting scheme can be applied for controlling the variability of factor weights (see e.g., Saati and Memariani 2005, Omrani 2013). The earliest and also most cited approach goes back to Roll et al. (1991) and, therefore, is described in the following.

Roll et al. (1991) introduced three different techniques to obtain a common set of weights. For the sake of simplification, only their first idea is explained here. This approach requires the computation of a basic DEA model (e.g., the model proposed by Charnes et al. 1978) in a first step. In a second step, the arithmetic average of the weights for each input and output factor is determined and, subsequently, used to compute the DMUs’ efficiency scores. Mathematically, this second step can be expressed for a DMU p as follows:

$$Eff\,(X_p, Y_p) = \sum_{r=1}^{s} \bar{\mu}_r \, y_{rp} \Big/ \sum_{i=1}^{m} \bar{v}_i \, x_{ip}, \qquad \forall j \tag{4.1}$$

whereby $\bar{v}_i$ and $\bar{\mu}_r$ are the arithmetic averages of the optimal weights received from the programming problem (3.4) (i.e., $\bar{v}_i = 1/n \sum_{j=1}^{n} v_{ij}^*, \ \forall i$ and $\bar{\mu}_r = 1/n \sum_{j=1}^{n} \mu_{rj}^*, \ \forall r$).

The major advantage of this approach is its ease of application. There is no further mathematical programming necessary as it can be computed using basic DEA solvers.[51] However, one needs to recall that such mean weights may also lead to infeasible solutions (see Roll et al. 1991, p. 7).

Since the approach is based on basic DEA models (in a first step) and a common set of weights to compute the DMUs' efficiency scores (in a second step), one can clearly argue that the approach simultaneously incorporates characteristics of decentralized and centralized management scenarios. Hence, it is straightforward to classify the method of Roll et al. (1991) as a hybrid DEA approach. Other hybrid and centralized DEA approaches for controlling factor weights are shown in Table A1 of the Appendix.

4.4.2.3 Classification schemes

In many practical situations, organizations need to group a set of DMUs according to different performance indicators. For example, central management may be required to predict whether a subordinated business unit is likely to default or not when the economic situation deteriorates (see Sueyoshi 1999). Therefore, the organization may consider different performance criteria (e.g., the perceived product quality or the staff's educational level) and, subsequently, allocate the DMUs to one of the two groups "default of business unit is likely" and "default of business unit is not likely" (see Sueyoshi 1999). Since traditional business accounting approaches such as ABC analysis usually fail to simultaneously incorporate multiple decision criteria in a plausible manner, a couple of DEA-based approaches have been suggested in recent decades. They can be divided into two categories:

The *first category* assumes that there are no predefined groups for any observation and, hence, the entire unit set needs to be categorized. For example, Amirteimoori and Kordrostami (2013) introduced an approach that is able to cluster a set of DMUs according to their respective size. In accordance with the conventional ABC analysis, Chen (2011) developed a multi-step procedure for classifying observations into three different catego-

[51] See Barr (2004) for an overview of different DEA software packages and the available models.

ries (e.g., very important (group A), moderately important (group B), and relatively unimportant units (group C)), while simultaneously considering several distinct performance indicators. A similar approach has been suggested by Hatefi and Torabi (2015). However, their approach comprises only a single methodological step and, therefore, shows major computational advantages compared to the idea of Chen (2011).

The *second category* of DEA-based classification schemes assumes that some observations have been already allocated to a set of predefined groups. In that case, the respective approach needs to focus only on the correct allocation of the newly sampled DMUs, which requires a detailed analysis of any similarities with the predefined group members. The only approaches of this second category have been proposed by Sueyoshi (1999, 2001).

The approach of Sueyoshi (1999) has received the most citations from other hybrid and centralized DEA approaches. Therefore, his idea is described in more detail below. An overview of all DEA-based classification schemes, which either assume a hybrid or centralized management scenario, are shown in Table A2 in the Appendix.

The approach of Sueyoshi (1999) combines two different techniques – DEA and Discriminant Analysis (DA) – and consists of the subsequent solution of two linear programming problems. Each programming problem yields an evaluation score that can be used for the determination of the group membership of a newly sampled DMU.

Assume that there are two predefined groups denoted as $g = 1$ and $g = 2$ and that the respective (predefined) quantity of observations in each category is represented through δ_1 and δ_2, respectively. In addition, suppose that $\delta_1 + \delta_2 = n$ is satisfied and that the performance of DMU j regarding indicator b ($b = 1,\ldots,R$) is denoted as I_{bj}. Considering this notation, the mathematical programming problem of the *first step* of the approach can be formulated as follows:

$$
\begin{aligned}
&\min \sum_{j=1}^{\delta_1} S^{+}_{(g=1)j} + \sum_{j=1}^{\delta_2} S^{-}_{(g=2)j} \\
&s.t. \qquad \sum_{b=1}^{R} \alpha_b I_{bj} + S^{+}_{(g=1)j} - S^{-}_{(g=1)j} = \sigma, \qquad \forall j \in g = 1 \\
&\qquad\qquad \sum_{b=1}^{R} \beta_b I_{bj} + S^{+}_{(g=2)j} - S^{-}_{(g=2)j} = \sigma - \eta, \qquad \forall j \in g = 2 \\
&\qquad\qquad \sum_{b=1}^{R} \alpha_b = 1 \\
&\qquad\qquad \sum_{b=1}^{R} \beta_b = 1 \\
&\qquad\qquad S^{+}_{gj}, S^{-}_{gj}, \alpha_b, \beta_b \geq 0, \qquad \forall b, \forall g, \forall j
\end{aligned}
\tag{4.2}
$$

where η is a small number to impose a minimal gap between the threshold value σ of the two groups and avoids trivial solutions (e.g., Sueyoshi 1999, p. 566). The weights attached to the respective factor I_{bj} are denoted by α_b and β_b, respectively. The variables $S^{+}_{(g=1)j}$ and $S^{-}_{(g=2)j}$ represent slack-based distance parameters, which indicate how much $\sum_{b=1}^{R} \alpha_b I_{bj}$ and $\sum_{b=1}^{R} \beta_b I_{bj}$ are separated from the threshold score σ. Similar to the additive DEA model of Charnes et al. (1985), the objective function of (4.2) seeks to minimize these two slack-based distance measures.

Note that (4.2) uses group specific weights (e.g., α_b and β_b) to generate two discriminant functions that separate the groups in the multidimensional space. Through the application of the optimal factor weights, denoted in the following as α^*_b and β^*_b, one can compute the weighted performance score for a newly sampled observation *p* as $\sum_{b=1}^{R} \alpha^*_b I_{bp}$ and $\sum_{b=1}^{R} \beta^*_b I_{bp}$, respectively. Comparing these values with the *optimal* threshold value σ^* indicates whether the newly sampled observation belongs (a) to group g = 1, (b) to group g = 2 or (c) to the overlap of the two discriminant functions, which is denoted in the following as $G_1 \cap G_2$. Hence, the first programming problem tests whether a newly sampled DMU can be *unambiguously* allocated to one of the two groups or is enveloped by the overlap of the discriminant functions $G_1 \cap G_2$.

The *second step* of the approach then determines a single discrimination function that allows classifying whether the observations belonging to the overlap $G_1 \cap G_2$ should be rather classified to group g = 1 or g = 2, respectively (see Sueyoshi 1999). Since a common set of weights is applied during this stage, the corresponding programming problems implicitly assume a centralized management scenario. However, the first stage of this approach uses separate sets of weights for each group, which means that the entire framework needs to be classified as a hybrid DEA approach.

A major benefit of the approach of Sueyoshi (1999) is that it is based on less restrictive assumptions than conventional DA. For example, researchers that use DA usually need to estimate the underlying distribution (e.g., normal distribution) for the data set of the two groups. However, it is well known that many practical data sets do not satisfy such theoretically expected distributions. Therefore, many results of a DA are highly questionable (see Sueyoshi 2001, p. 328).

A major deficit of the aforementioned approach is its limitation to positive data sets. This substantially reduces its applicability for financial data where negative values are typical, e.g., to express net losses. Furthermore, the use of multiple discriminant functions (two functions in the first programming problem and one function in the second programming problem) reduces the computational performance of this approach. This drawback is especially crucial when the concept may be applied to large scale simulation data (see Sueyoshi 2001, p. 330).

4.4.2.4 Resource allocation and target setting

In order to improve the overall performance of an organization, a central decision maker frequently seeks to set targets and allocate a permitted level of costs to a group of subordinated operating units. The allocation of a university's overhead costs to a number of different scientific departments (see e.g., Beasley 2003, p. 198) or the distribution of financial grants to local authorities (see e.g., Athanassopoulos 1995, p. 542) are some examples where such allocation problems occur. In the majority of cases, a mixture of business accounting instruments and negotiations are applied to distribute resources and set targets in a plausible manner (see Beasley 2003, pp. 198-199). However, the application of these instruments is challenging when one needs to simultaneously incorporate multiple allocation criteria (e.g., different cost drivers), which cannot be simply aggregated on a monetary or other basis.

To address this issue, the literature on performance measurement comprises a number of different hybrid or centralized DEA approaches (see e.g., Thanassoulis 1996, Amirteimoori and Kordrostami 2005, Bi et al. 2011, Li et al. 2017). Since the approach of Beasley (2003) has received the most citations within this research stream, it is described in more detail below. The remaining hybrid or centralized DEA approaches are listed in Table A3 of the Appendix.

The approach of Beasley (2003) is originally designed to allocate a fixed quantity of overhead costs to a set of operating units. The entire approach comprises five distinct methodological steps, which need to be conducted in an iterative manner. Four different mathematical programming problems are applied in Step 1–4. For the sake of simplification, the focus of the following description lies on the first of the four programming problems, which is also most suitable to show the fundamental idea behind this resource allocation approach.[52]

Suppose that the entire amount of fixed overhead costs to be allocated is denoted as *F* and that the respective quantity allocated to DMU *j* is mathematically expressed through

---

[52] For a more thoroughly explanation, the reader is referred to Beasley (2003).

f_j (j = 1,…,n). In addition, assume that $F = \sum_{j=1}^{n} f_j$ and $f_j \geq 0 \; \forall \; j = 1,...,n$ are consistently satisfied. In this case, the first programming problem of Beasley (2003) can be mathematically reformulated as follows:[53]

$$\begin{aligned}
&\max \bar{\theta} \\
&s.t. \quad && \bar{\theta} = \sum_{j=1}^{n} \theta_j / n \\
& && \theta_j = \sum_{r=1}^{s} \mu_r y_{rj} \Bigg/ \left(\sum_{i=1}^{m} \nu_i x_{ij} + f_j \right), \; \forall j \\
& && F = \sum_{j=1}^{n} f_j \qquad\qquad (4.3) \\
& && f_j \geq 0, \qquad \forall j \\
& && 0 \leq \theta_j \leq 1, \qquad \forall j \\
& && \mu_r \geq \varepsilon, \qquad \forall r \\
& && \nu_i \geq \varepsilon, \qquad \forall i.
\end{aligned}$$

Constraint 2 of (4.3) represents the efficiency of each DMU j (j = 1,…,n). However, in contrast to the usual DEA definition of efficiency (see programming problem (3.4)), formula (4.3) attaches a predetermined weight of one to the allocated cost quantity f_j. The objective function in combination with constraint 1 indicates that the approach seeks for a cost allocation, which enables a maximum average efficiency $\bar{\theta}$ while applying identical weights to each DMU (i.e., a common set of weights). In this way, (4.3) aims at a cost distribution that is considered by each DMU as fair and equitable (see Beasley 2003, p. 202). In the extreme case, the allocation of overhead costs allows each DMU to obtain an efficiency score of up to 100 %.

Because of the common set of weights that is applied to each observation under consideration, the first programming problem is implicitly based on the centralized management

53 Note that a simplified version of the approach of Beasley (2003) is described here. For example, it has been neglected that there may be a set of DMUs whose received cost quantity has been fixed in advance. This aspect is considered in the original approach (see Beasley 2003, p. 206).

assumption. That is, the entire cost allocation is conducted in strict consideration of the particular preferences of the central decision maker and, therefore, only provides a minimum degree of decision making authority to the subordinated DMUs. However, since DMU specific weights are also applied in subsequent steps of the methodology,[54] the approach needs to be classified as hybrid.

The approach presented in (4.3) has generalized previous frameworks (see e.g., Golany et al. 1993, Golany and Tamir 1995, Thanassoulis 1996, Thanassoulis 1998) and, therefore, enlarged the potential application area of DEA-based allocation approaches. It has also provided the flexibility to incorporate further judgements and preferences of the central decision maker through additional constraints (e.g., weight restrictions). This is usually not possible when standard accounting instruments are applied (see Beasley 2003, p. 199). However, the number of different non-linear programming problems as well as the iterative methodology makes the application of this approach mathematically complex and inconvenient in practice. This substantial drawback has been solved by simplified DEA models proposed by, e.g., Lozano and Villa (2004) or Mar-Molinero et al. (2014). These approaches are discussed in more detail in Section 6.3.

4.4.2.5 Ranking of DMUs

In practice, one may encounter situations where a complete and consistent ranking of all DMUs is required. For example, governmental financial spending for a university may be strictly bound to the university's respective position on a nation-wide science performance ranking. The literature comprises a considerable amount of corresponding DEA-based approaches, which seek to obtain a consistent ranking for efficient as well as inefficient DMUs (see e.g., Liu and Peng 2008, Jahanshahloo et al. 2010).[55] In total, 40 different hybrid or centralized DEA approaches were identified, of which an overview is provided in Table A4 in the Appendix. As the most cited approach has been published by Kao and Hung (2005), it is introduced in more detail in the following.

---

[54] See Step 2 of the approach of Beasley (2003).

[55] See Adler et al. (2002) for a comprehensive review of different DEA-based ranking methods.

The approach of Kao and Hung (2005) is based on the subsequent solution of two distinct mathematical programming problems: In a *first step*, a basic DEA model (i.e. CCR model) is solved to obtain a set of optimal performance scores for each DMU *j*. This set of values is denoted in the following as $\theta^* = (\theta_1^*, ..., \theta_n^*)$. In a *second step*, Kao and Hung (2005) apply a programming problem, which yields new performance scores as close as possible to the performance scores received from the previous step. Let $\theta^{CSW*} = (\theta_1^{CSW*}, ..., \theta_n^{CSW*})$ denote the respective values from the second programming problem, whereby the superscript *CSW* indicates that these performance scores have been obtained applying a common set of weights. The entire mathematical programming problem of the second step of the approach of Kao and Hung (2005) is presented below:

$$\begin{aligned}
&\min \left[\sum_{j=1}^{n} (\theta_j^* - \theta_j^{CSW})^{\rho} \right]^{1/\rho} \\
&s.t. \quad \theta_j^{CSW} = \sum_{r=1}^{s} \mu_r y_{rj} \Big/ \sum_{i=1}^{m} \nu_i x_{ij}, && \forall j \\
&\qquad\ \ \theta_j^{CSW} \le 1, && \forall j \\
&\qquad\ \ \mu_r \ge \varepsilon, && \forall r \\
&\qquad\ \ \nu_i \ge \varepsilon, && \forall i.
\end{aligned} \tag{4.4}$$

The superscript ρ in the objective function is a distance parameter that can be used to compute the respective deviations between θ_j^* and θ_j^{CSW} based on different distance norms. For example, a value of ρ = 1 means that the difference is measured according to the so-called Manhattan distance metric. This implies that each deviation $\theta_j^* - \theta_j^{CSW}$ is equally weighted. By contrast, an increasing value for the distance parameter ρ puts more weight on the larger deviations. In the most extreme case (i.e., $\rho = \infty$), the entire weight is put on the largest deviation (i.e., $\max\{\theta_1^* - \theta_1^{CSW}, ..., \theta_n^* - \theta_n^{CSW}\}$). Hence, the objective of the programming problem (4.4) would be identical to minimizing the maximum deviation between θ_j^* and θ_j^{CSW} (see Kao and Hung 2005, pp. 1198-1199).[56]

Constraint 1 in (4.4) defines the efficiency of each DMU as the ratio of the weighted sum of outputs to the weighted sum of inputs, which is a standard assumption of basic DEA models. Constraint 2 guarantees that the respective performance score of each DMU *j* is bounded to a value between 0 and 100 %. Constraints 3 and 4 show that the objective values of the second programming problem are received based on a common set of weights.

The mathematical programming problem given in (4.4) allows a consistent performance ranking of all DMUs (i.e., efficient and inefficient ones) and, in addition, provides better comparability since the performance scores are computed on a common basis. However, it is important to note that even with the results received from (4.4), multiple DMUs can share the same rank. Therefore, it might also be possible that several DMUs are placed at position 1.[57]

Exclusively considering the programming problem presented in (4.4), one could straightforwardly argue that the approach is designed in correspondence with the idea of a centralized management scenario. However, since the approach requires the ex-ante solution of basic DEA models in order to obtain $\theta^* = (\theta_1^*, ..., \theta_n^*)$, the framework does not consistently apply a common set of weights to all DMUs. Instead, the weights are derived in a hybrid way, i.e. the weights are calculated with the connection to the DMUs' individual weights in the first DEA model. Therefore (and in line with our definitions given in Section 4.3), the publication of Kao and Hung (2005) is classified as a hybrid DEA approach.

---

56 The idea of using different distance norm in the context of operations research has been discussed by Dyckhoff (1985). For a discussion on different distance metrics and their respective application in the field of performance measurement, the interested reader is referred to Kleine and Glaser (2004) as well as Ahn et al. (2007). See Wang (2006) for a general discussion on different distance metrics from a mathematical perspective.

57 See the empirical illustration of Kao and Hung (2005, p. 1200).

4.4.2.6 Improving the discrimination power

As it has been described in Section 3.2.4, one of the major drawbacks of basic DEA models is their poor discrimination power regarding performance scores. Especially when a greater number of input and output factors are applied to evaluate a small number of observations, a large proportion of DMUs is classified as efficient. The reason is that the more input or output factors are included in the DEA, the more possibilities every DMU has to obtain an efficiency score of 100 % via the advantageous combination of its individual factor weights.

Similar to the approaches seeking to avoid extreme weights (see Section 4.4.2.2), the improvement of the discrimination power can also be achieved using weight restrictions. Corresponding approaches have been published by numerous authors (see e.g., Allen et al. 1997 as well as Dyson and Thanassoulis 1988). However, only little methodological support regarding this topic comes from approaches that are based on common weighting schemes: Only three different publications propose a hybrid or centralized DEA approach to improve the discrimination power between evaluated DMUs. The most cited approach has been published by Karsak and Ahiska (2005) and, hence, is described below. Table A5 of the Appendix also comprises the two alternative DEA approaches that apply a common set of weights.

The approach of Karsak and Ahiska (2005) is based on a mathematical programming problem which simultaneously computes for each DMU *j* the difference d_j to the ideal efficiency of 100 %. Therefore, one can determine the efficiency of an arbitrary chosen DMU *j* by simply calculating $1-d_j$. The complete mathematical formulation of the corresponding approach is given below:

$$
\begin{aligned}
&\min M \\
&s.t. \quad M \geq d_j, && \forall j \\
&\qquad \sum_{r=1}^{s} \mu_r y_{rj} \Big/ x_j + d_j = 1, && \forall j \qquad (4.5) \\
&\qquad d_j \geq 0, && \forall j \\
&\qquad \mu_r \geq \varepsilon, && \forall r.
\end{aligned}
$$

The variable *M*, which is minimized by the objective function of (4.5), represents the maximum deviation from efficiency (i.e., $M = \max\{d_1,...,d_n\}$). Therefore, Karsak and Ahiska (2005) refer to this variable also as the "minimax efficiency measure". Since the approach simultaneously evaluates all observations, the programming problem does not need to be solved for each DMU separately, which is a substantial computational advantage compared to basic DEA models. However, a major drawback of the approach is its limitation to the single input multiple output case. As indicated by constraints 2 and 4 in (4.5), there is only one subscript attached to input x_j. Therefore, the respective researcher is not allowed to incorporate *m* different input variables as in basic DEA models (compare constraint 2 in (3.3)).[58]

The variable μ_r ($r = 1,\ldots,s$) represents the weight of output *r* which is jointly applied to all DMUs. Because of this severe restriction compared to basic DEA models, the DMUs are not able to freely choose factor weights according to their individual preferences. Consequently, the probability of achieving an efficiency score of 100 % is significantly reduced for each DMU, which in turn improves the discrimination power compared to unrestricted (i.e., basic) DEA approaches. Due to the common set of weights applied for this purpose, it is clear that the publication of Karsak and Ahiska (2005) needs to be classified as a centralized DEA approach.

4.4.2.7 Finding the (single) most efficient DMU

The basic DEA models usually identify a *set* of several efficient DMUs depending on, e.g., the ratio of applied input-output factors, the number of evaluated units and the homogeneity of the data set. However, in many practical cases, it may be important to identify a *single* most efficient DMU. For example, when the central management needs to choose between a bunch of different investment opportunities (e.g., due to the lack of sufficient financial resources), it may be necessary to determine a single best performing alternative.

---

[58] Note that one can reformulate the mathematical model for the multiple input single output case by attaching weights to the input factors in the third and last constraint of (4.5).

For these cases, a number of different extensions to the basic DEA models have been proposed. Some of these approaches are based on common weighting schemes (e.g., Cook and Kress 1990, Foroughi 2011a) whereby the most cited approach has been published by Amin and Toloo (2007). However, as already noted in Table 4.2, the mathematically corrected version introduced by Amin (2009) is described here. Other approaches allocated to the same research stream are shown in Table A6 of the Appendix.

The approach of Amin (2009) builds on the idea of Ertay et al. (2006) that aims at identifying the best facility layout design in manufacturing systems. However, whereas the linear programming problem of Ertay et al. (2006) needs to be solved *n* different times (once for each DMU) and uses DMU specific weights, the approach of Amin (2009) requires the solution of a single linear programming problem and applies a common set of weights to all DMUs. The corresponding mathematical programming problem is as follows:

$$
\begin{aligned}
&\min M \\
&s.t. \quad M \geq d_j, && \forall j \\
&\sum_{i=1}^{m} \nu_i x_{ij} \leq 1, && \forall j \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} + d_j = 0, && \forall j \\
&\sum_{j=1}^{n} \kappa_j = n-1 && \\
&\kappa_j - d_j \gamma_j = 0, && \forall j \\
&d_j \geq 0, \quad \gamma_j \geq 1, \quad \kappa_j \in \{0,1\}, && \forall j \\
&\nu_i \geq \varepsilon^*, && \forall i \\
&\mu_r \geq \varepsilon^*, && \forall r
\end{aligned}
\tag{4.6}
$$

whereby ε^* denotes the maximum non-Archimedean (see Amin and Toloo 2007, p. 74).[59]

[59] The idea of the maximum non-Archimedean has been originally suggested by Cook et al. (1996).

Similar to (4.5) and (4.7), the objective function in (4.6) also seeks to minimize the value of the variable *M* which represents the maximum deviation from efficiency. However, the mathematical programming problem presented in (4.6) is also applicable to multiple input multiple output scenarios (see constraints 2 and 3). Another substantial difference is that the mathematical restriction $\kappa_j - d_j \gamma_j = 0$ (see constraint 5) in combination with the requirement that the binary variable κ_j is always equal to $n-1$ (see constraint 4) ensures that there is only one DMU for which the difference from efficiency d_j is zero. The corresponding DMU is therefore considered as the most efficient DMU.[60]

As the third constraint of (4.6) shows, the programming problem applies the factor weights in a uniform way to the entire unit set. Hence, this approach implicitly assumes a centralized management scenario to identify the single most efficient DMU.

4.4.2.8 Construction of composite indicators

So-called composite indicators are a performance measurement tool where *various* distinct performance indicators are aggregated to generate a *single* performance score. This aggregation is frequently based on weights received from experts' judgements and, therefore, subjective. When experts' judgements are not available, it is often argued that equal weights can be used (e.g., the application of an arithmetic average) (see Zhou et al. 2007).

The application of composite indicators is especially popular in macroeconomics and politics to express countries' level of development. For example, the so-called human development index, which has been introduced by the United Nations, is computed as the geometric mean of three different performance indicators: the life expectancy index, the education index and the income index (see United Nations Development Program 2018). Another popular composite indicator is the environmental performance index that – based on value judgements – aggregates a total of 24 performance indicators to evaluate "how close countries are to established environmental policy goals" (see Yale University Center for Environmental Law & Policy 2018).

---

[60] Recall that the efficiency value can be calculated as 1 – d_j which is equal to an efficiency score of 100 % when $d_j = 0$ is satisfied.

The methodological aggregation of performance indicators based on value judgements has often been a major focus of criticism and research. Even the application of equal weights (as used by the human development index) can be considered as subjective. Therefore, Nardo et al. (2005) have proposed a number of different approaches for aggregating performance indicators. In line with the argumentation of Nardo et al. (2005), numerous authors have also introduced DEA approaches to construct composite indicators (see e.g., Domínguez-Serrano and Blancas 2011, Tofallis 2013). Thereby, the most cited approach, which is based on a common set of weights, goes back to Hatefi and Torabi (2010) and, hence, is thoroughly described below. Other hybrid or centralized DEA approaches that seek to aggregate various performance indicators to provide a single performance score are shown in Table A7 in the Appendix.

The approach of Hatefi and Torabi (2010) is based largely on the programming problem (4.5) (see Section 4.4.2.3) and applies a common set of weights "to enable a fair comparison" among all units under evaluation (see Hatefi and Torabi 2010, p. 116). Their suggested mathematical programming problem is as follows:

$$
\begin{aligned}
&\min M \\
&s.t. \quad M \geq d_j, && \forall j \\
&\qquad \sum_{b=1}^{R} \alpha_b I_{bj} + d_j = 1, && \forall j \\
&\qquad d_j \geq 0, && \forall j \\
&\qquad \alpha_b \geq \varepsilon, && \forall b.
\end{aligned}
\tag{4.7}
$$

From a comparison of (4.7) with the approach presented in (4.5), one can see that the objective function as well as constraints 1 and 3 are identical. The interpretation and the corresponding implications are similar to what has been given above; hence, they are not repeated here. The only substantial difference between (4.5) and (4.7) can be observed for constraint 2: The approach in (4.7) does not distinguish between input and output factors. Hence, it can be considered as a simplification of (4.5).

This simplification is possible because (4.7) only considers so-called benefit type performance indicators which are denoted as I_{bj}. This particular indicator type satisfies the property "the larger the better" (see Hatefi and Torabi 2010, p. 115) and, therefore, is comparable to the traditional definition of output factors. For so-called cost type performance indicators (i.e., indicators that do not fulfil the "the larger the better" property), Hatefi and Torabi (2010) suggest that they should be converted into benefit type performance indicators by using their reciprocal values.

As described before, the different performance indicators are aggregated in (4.7) via the application of a common set of weights. The respective weight attached to performance indicator b is denoted as α_b in (4.7). Therefore, one could compute the value of the constructed composite indicator of DMU p as $\sum_{b=1}^{R}\alpha_b^* I_{bp}$ whereby α_b^* denotes the optimal set of common weights received from the solution of (4.7). Since Hatefi and Torabi (2010) exclusively apply a common set of weights to all DMUs, it is straightforward to categorize their approach as a centralized DEA model.

4.4.2.9 Dynamic performance measurement

In many practical cases, it is necessary to compare a DMU's performance not only in a *static* setting, but also over time and, therefore, in a *dynamic* context. Such dynamic performance comparisons can be conducted by means of the previously discussed Malmquist productivity index (see Section 3.4.2) or, alternatively, via the application of the so-called window analysis (see e.g., Webb 2003). These approaches have in common that they are based on basic DEA models. Therefore, the respective performance scores are computed regarding technologies at different time periods using the most favorable set of weights for each DMU. In most cases, these weights differ widely and, hence, may not allow an equitable and consistent performance comparison between units – especially if multiple time periods are involved (see Kao 2010). Furthermore, the application of these approaches is not appropriate if the focal organization is characterized by a high degree of centralization.

Despite these substantial limitations of the conventional dynamic performance measurement approaches, there is only little methodological support regarding this topic. Only three publications – namely Kao (2010), Yang et al. (2016) as well as Afsharian and Ahn (2017) – apply a common set of weights for measuring a DMU's productivity changes over time and, therefore, can be considered either as a hybrid or centralized DEA approach. The most cited approach was published by Kao (2010) and, hence, is briefly presented below. Table A8 in the Appendix of this publication provides an overview of all identified (hybrid and centralized) DEA approaches allocated to the research stream *Dynamic performance measurement*.

The approach of Kao (2010) comprises three methodological steps: In a *first step*, each observation is evaluated using the metafrontier DEA approach given in (3.11). Hence, the different DMUs are allowed to choose the most favorable set of weights and are compared to a global benchmark technology that is built as a convex combination of all observations in all time periods. In a *second step*, Kao (2010) proposes to compute a set of objective values as close as possible to the performance scores obtained from Step 1. Hence, he suggests a slightly modified version of the programming problem presented in (4.4). Based on the common set of weights that has been received through the solution of (4.4), the *third step* computes the efficiency scores for all n DMUs (j = 1,…,n) observed in all t (t = 1,…T) and, straightforwardly, applies the following Malmquist index to measure their respective productivity changes over time:

$$MI_p^{CSW} = \theta_p^{t+1,CSW*} / \theta_p^{t,CSW*} \tag{4.8}$$

whereby $\theta_p^{t,CSW*}$ and $\theta_p^{t+1,CSW*}$ denote the computed efficiency scores of DMU p for the time periods t and t+1, respectively. Again, the superscript *CSW* indicates that the respective performance values are computed via the application of a common set of weights (based on (4.4)).

Identical to the economic interpretation of the conventional Malmquist index, a MI_p^{CSW} value above unity corresponds to a DMU's productivity improvement from period t to t+1. A value below unity indicates the opposite (i.e., productivity deterioration). If the respective value of MI_p^{CSW} is equal to one, identical productivity levels exist in period t and t+1.

The index in (4.8) is a special case of the global Malmquist productivity index proposed by Pastor and Lovell (2005) and, hence, possesses all its properties. That is, the index satisfies the circularity property and is immune to infeasibilities in the presence of VRS.

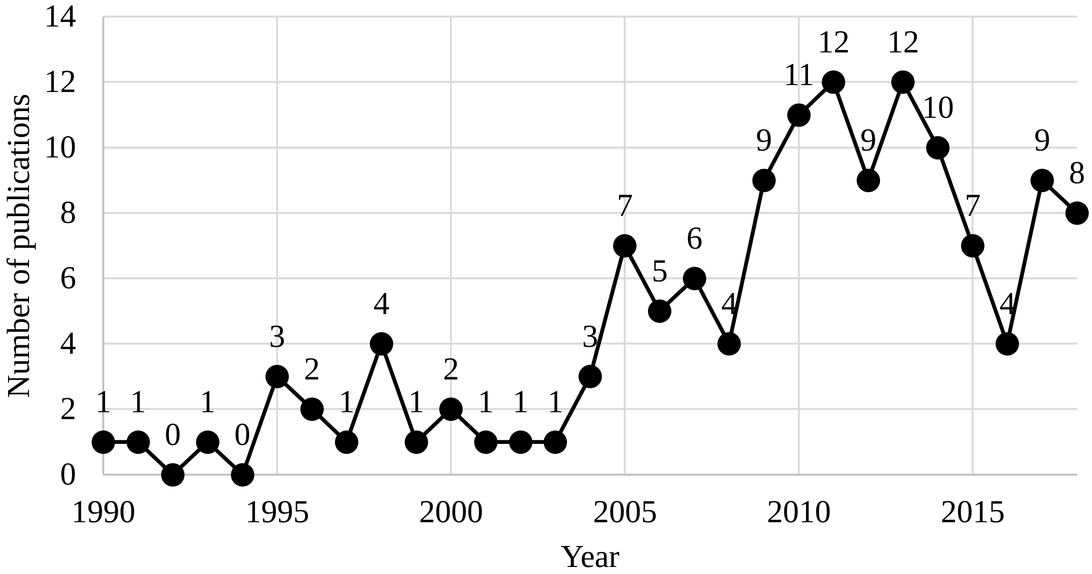
However, one additional property which is not shared with the global Malmquist productivity index is that the efficiency of every DMU is computed using a common set of weights that makes the calculated performance indices comparable among all DMUs (see Kao 2010).

Note that the Malmquist index in (4.8) is based on the strong premise that the technology remains unchanged between the start and the end of the analysis (see Afsharian and Ahn 2017). Accordingly, it is assumed that neither the external environment (such as government rules and regulations as well as the economic condition) nor the internal environment (such as the organizational strategies, internal rules and regulations and policy directives) has changed over the periods of time under consideration. However, this assumption is not likely to be satisfied in most real applications. Consequently, including all convex combinations of all observations in all time periods with different technologies in the analysis may by inappropriate (see Afsharian and Ahn 2017). Because of this, Afsharian and Ahn (2017) proposed an extension of (4.8). They suggested applying a Malmquist productivity index, which makes use of a non-convex global benchmark technology and, therefore, preserves the individual characteristics of the contemporaneous technologies over time which can be traced later in measuring productivity changes.

4.4.3 Findings and synopsis

In total, 135 hybrid or centralized DEA approaches were identified through the applied research methodology (see Table 4.1 in Section 4.4.2.1). Figure 4.4 shows how many articles were published in each year from 1990 until 2018 and the corresponding development can be separated into three steps. In a first phase, from 1990 until 2003, the number of publications was low. But beginning from 2004, a significant growth of publications can be observed, with a peak of 12 papers in 2011 and 2013, respectively. Thereafter, a slight decline in publishing took place, leading to only eight papers in 2018.

Figure 4.4: Distribution of publications by year (1990–2018)



The development in the number of published articles resembles a typical product lifecycle. However, the indicated diminishing interest for modeling different degrees of centralization is counterintuitive, since there exists a huge variety of organizational settings in which the DMUs to be evaluated are at least partially under the influence of a central decision maker. This especially includes all cases where DMUs are – or can be – incentivized depending on their performance. Given the importance of such scenarios in practice, the present review may increase the awareness of the research field and motivate researchers to contribute.

Table 4.3 shows the top 17 journals that have published the most hybrid or centralized DEA approaches between 1990 and 2018. The most articles were published in the *European Journal of Operational Research*. On the second and third rank are *Expert Systems with Applications* and *Journal of the Operational Research Society*, respectively. The journals *Omega* and *Computers & Industrial Engineering* share the fourth position.[61] Altogether, these five journals published more than one third of the hybrid or centralized DEA approaches. Considering the scope of the journals, the finding seems reasonable because the application areas of DEA can be mainly related to the field of operations research which is in line with Emrouznejad et al. (2008) as well as Emrouznejad and Yang

[61] With the exception of *Expert Systems with Applications*, the other four journals are identical with the top journals identified by Emrouznejad and Yang (2018).

(2018).[62] In this respect, the identified hybrid and centralized approaches do not show substantial particularities compared to DEA-related research in general. One reason for this may be the heterogeneous application areas of common weighting schemes as indicated by the different research streams discussed in Section 4.4.2.

Table 4.3: The 17 journals with the most publications

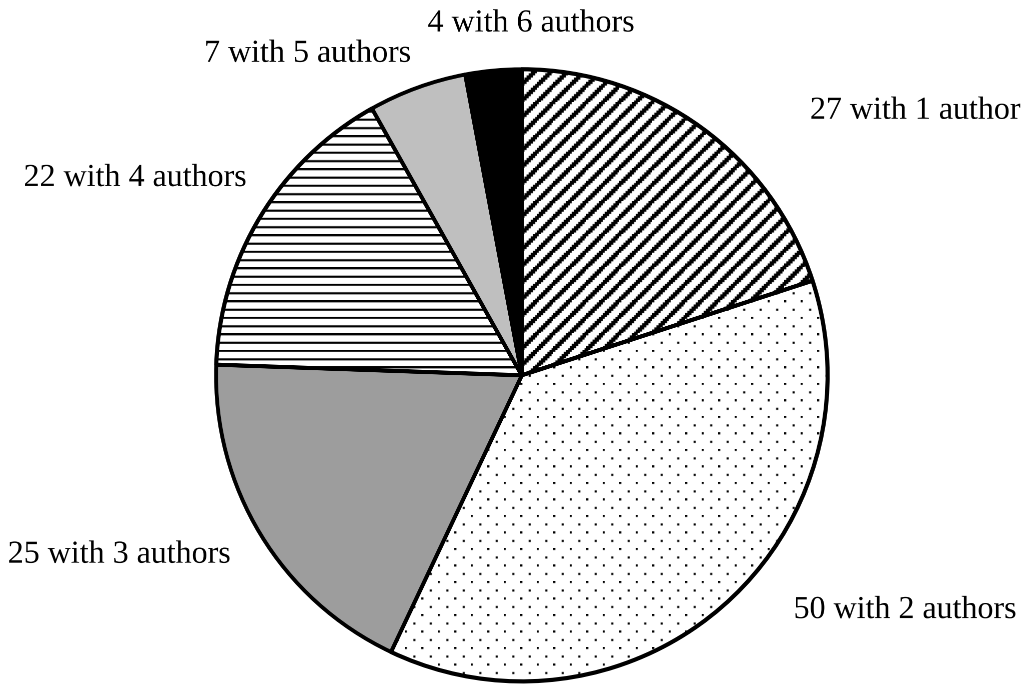
Name of journal	Number of publications	Percentage Share	Cumulated percentage share
1 European Journal of Operational Research	12	8.89 %	8.89 %
2 Expert Systems with Applications	11	8.15 %	17.04 %
3 Journal of the Operational Research Society	10	7.41 %	24.44 %
4 Omega	9	6.67 %	31.11 %
4 Computers & Industrial Engineering	9	6.67 %	37.78 %
6 Applied Mathematical Modelling	6	4.44 %	42.22 %
6 Applied Mathematics and Computations	6	4.44 %	46.67 %
7 International Journal of Production Research	5	3.70 %	50.37 %
8 Journal of Productivity Analysis	4	2.96 %	53.33 %
9 Social Indicators Research	3	2.22 %	55.56 %
9 Computers and Operations Research	3	2.22 %	57.78 %
9 Annals of Operations Research	3	2.22 %	60.00 %
9 Central European Journal of Operations Research	3	2.22 %	62.22 %
13 Management Science	2	1.48 %	63.70 %
13 Measurement	2	1.48 %	65.19 %
13 Journal of Applied Mathematics	2	1.48 %	66.67 %
13 Asia-Pacific Journal of Operational Research	2	1.48 %	68.15 %
Total	**135**	**100 %**	**100 %**

The literature review showed that 208 different authors have made contributions to the field of hybrid or centralized DEA. Figure 4.5 provides an overview of the distribution of the number of authors per publication. Of the 135 identified publications, 27 articles (20.0 %) were published by a single author. The largest share of publications (50 publications or 37.0 %, respectively) was written by two authors. The remaining 43.0 % were developed by three or more authors. Four publications were written by six different authors. The corresponding average per publication is 2.6 which is identical to the corresponding value calculated by Emrouznejad and Yang (2018) for the entire field of DEA.

[62] Emrouznejad et al. (2008) and Emrouznejad and Yang (2018) studied the scholarly literature of DEA and – inter alia – analyzed the number of DEA-related publications per journal.

Hence, this result also does not suggest a substantial difference compared to the general DEA research area.

Figure 4.5: Distribution of the number of authors per publication



Based on their conceptual similarity, one can cluster the keywords of all publications into 21 distinct “keyword categories”. The top eleven categories are shown in Table 4.4. This table shows a high proportion of publications that is related to the keyword *DEA* (118 publications or 87.4 %, respectively). The second most frequently used category (50 publications or 37.0 %, respectively) comprises terms such as common weights, common set of weights or common weight analysis.

Many of the keywords listed in Table 4.4 can be directly related to a certain research stream of Section 4.4.2. For example, the keyword categories *Ranking, DEA Ranking* (25 publications or 18.5 %, respectively) and *Resource allocations, Fixed cost allocation, target setting* (27 publications or 20.0 %, respectively) can be assigned to the research streams *Ranking of DMUs* (see Section 4.4.2.5) and *Resource allocation and target setting* (see Section 4.4.2.4), respectively.[63]

Interestingly, only five publications used terms such as *Centralized Management* to categorize their articles. No publication was found that applied keywords with a similarity to the idea, considered here, of a hybrid management scenario or other organization theoretical aspects. Both findings are in line with the aforementioned production-oriented foundations of DEA and, hence, prove the importance of this systematic literature review as well as a fundamental research gap. Furthermore, the results indicate that the majority of researchers is probably not fully aware of their implicit centralized or hybrid management assumptions. This is crucial when the respective approaches are applied to highly decentralized organizations as the received performance scores may not appropriately reflect the underlying practical setting. Therefore, the economic interpretation as well as the corresponding conclusions may be flawed.

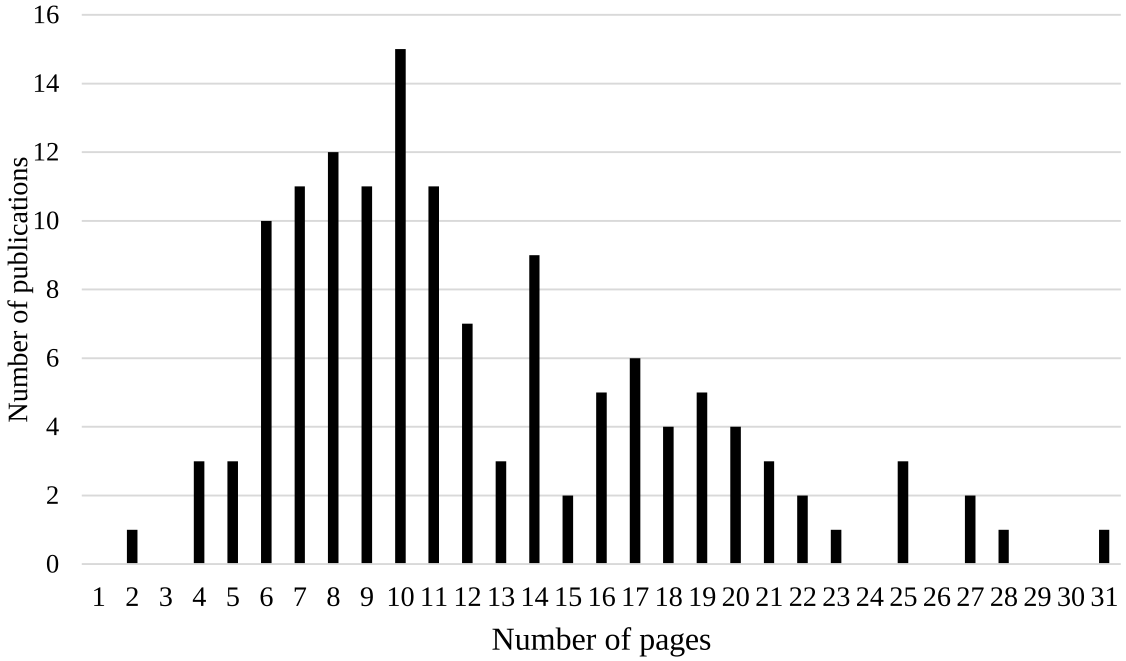
Table 4.4: The top eleven categories of keywords

Category	Number of publications
1 DEA, Integrated DEA, Fuzzy DEA, Robust DEA, Network DEA	118
2 Common weights, Common set of weights, Common weight analysis	50
3 Efficiency, Super efficiency, Absolute efficiency, Efficiency invariance	33
4 Resource allocation, Fixed cost allocation, Target setting	27
5 Ranking, DEA ranking	25
6 MCDA, MC, MCDM	10
7 Cluster analysis, ABC inventory classification	9
7 Most efficient DMU, Ideal DMU	9
9 Index, Composite indicators, Human development indicator	8
9 Multi objective optimization, Multiobjective programming, MOLP	8
11 Weight restrictions, Criteria weights, Virtual weights restriction	7
11 Cross efficiency, Cross efficiency evaluation	7

[63] Other research streams can also be assigned to the keyword categories in Table 4.4. The DEA-based *classification schemes* (see Section 4.4.2.3) apply to the category *Cluster analysis*, *ABC inventory classification*. The approaches for *Finding the (single) most efficient DMU* (see Section 4.4.2.7) correspond to the keyword category *Most efficient DMU*, *Ideal DMU*. Approaches that suggest the *Construction of composite indicators* (see Section 4.4.2.8) apply to the keyword category *Index, Composite indicators, Human Development Indicator*.

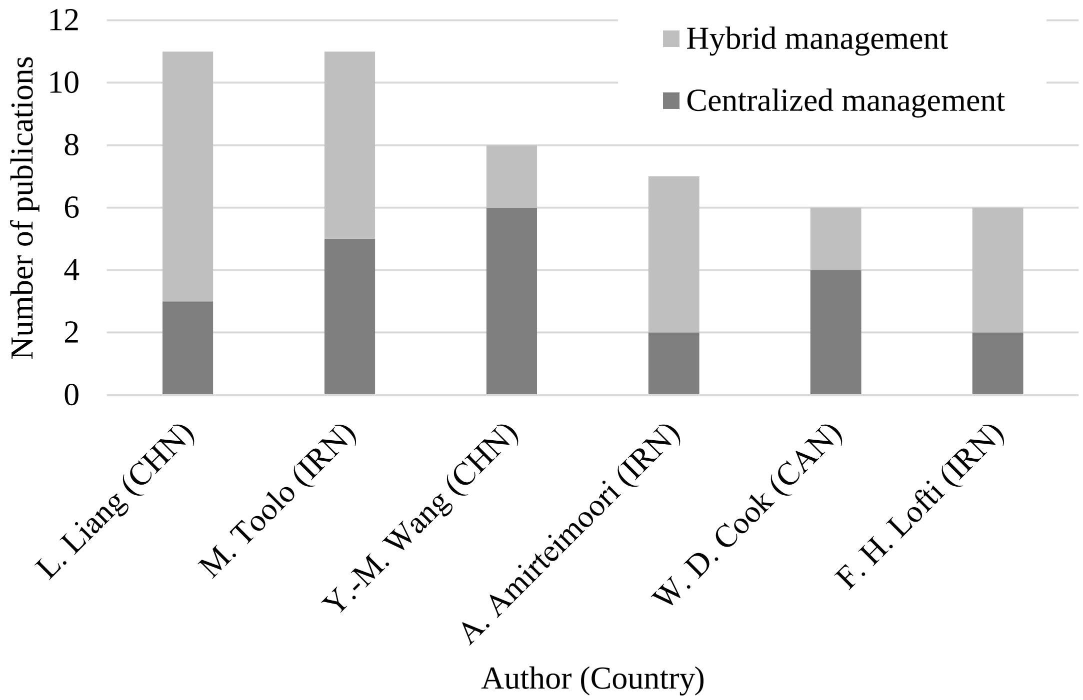
According to the data set used, 1,658 pages regarding hybrid or centralized DEA approaches were published since 1990. Figure 4.6 depicts the distribution of pages and one can see similarities with a skewed Gaussian distribution. While the longest article consists of 31 pages, the shortest publication only has two pages. The average length of the publication is around 12.3 pages which is again identical with the corresponding value of the study of Emrouznejad and Yang (2018).

Figure 4.6: Distribution by number of pages



The most productive authors (i.e., the authors who have published the highest number of papers about hybrid or centralized DEA approaches) are shown in Figure 4.6. L. Liang and M. Toloo published eleven approaches each and, therefore, are the authors with the most publications in this research area. Two researchers listed in Figure 4.7 are from China. The only American researcher listed in Figure 4.7: The authors with the most publications is Wade D. Cook (position 5 in our ranking) who was also identified by Emrouznejad et al. (2008) as one of the most productive authors in the entire DEA research field. Interestingly, three of the six most productive authors were researchers from Iran. Two of them (e.g., A. Amirteimoori and F. H. Lofti) were even employed at the same university (e.g., Ismalic Azad University).

Figure 4.7: The authors with the most publications



A comparison of the cumulated share of all authors with their respective cumulated share of the publications considered shows that the top 10 % of the most productive authors wrote 31.25 % of the articles (see Table 4.5). The top 20 % of the most productive authors published 47.28 % of all hybrid or centralized DEA approaches. This is a relatively small proportion if one compares this results with the statistics provided by Lee et al. (2014). They analyzed the development pattern of the DEA research field and have shown that almost 60 % of all publications were produced by only 20 % of the authors. A potential reason for this difference may be that there is only a small share of authors specializing on the research of common weighting schemes in the field of DEA. However, the review also demonstrates that common weight approaches contain attractive features that make them valuable for various economic situations. Therefore, it is only reasonable that the set of contributing authors is highly heterogeneous too.

Table 4.5: Cumulated share of publications by cumulated share of authors

Cumulated share of authors	Cumulated share of publications
10 %	31.25 %
20 %	47.28 %
30 %	58.70 %
40 %	66.30 %
50 %	72.01 %
60 %	77.45 %
70 %	83.15 %
80 %	88.86 %
90 %	94.57 %
100 %	100 %

When the number of published articles per author is plotted on the x-axis against the respective frequency (i.e., the number of authors who published a certain amount of hybrid or centralized DEA approaches) on the y-axis, one receives the distribution shown in Figure 4.8. The bar nearly at the right edge of Figure 4.8 depicts the case that an author published eleven articles. As it has been mentioned before, this was only achieved by two authors (namely L. Liang and M. Toloo). By contrast, 143 authors have published one hybrid or centralized DEA approach (the bar at the left end of Figure 4.8). When the respective natural log values of the two variables are plotted, one obtains the graph shown in Figure 4.9. The graph shows that the distribution of published articles per author follows a power law. That is, a very small number of authors published a high number of articles, but numerous authors have written a few (i.e., one or two) articles. This finding is in line with the aforementioned interpretation of Table 4.5 and proves the extensive heterogeneity of the DEA research field considered here. Furthermore, it can be concluded that only some authors specialize on the application of common weighting schemes.

Figure 4.8: Distribution of frequency by the number of publications per author

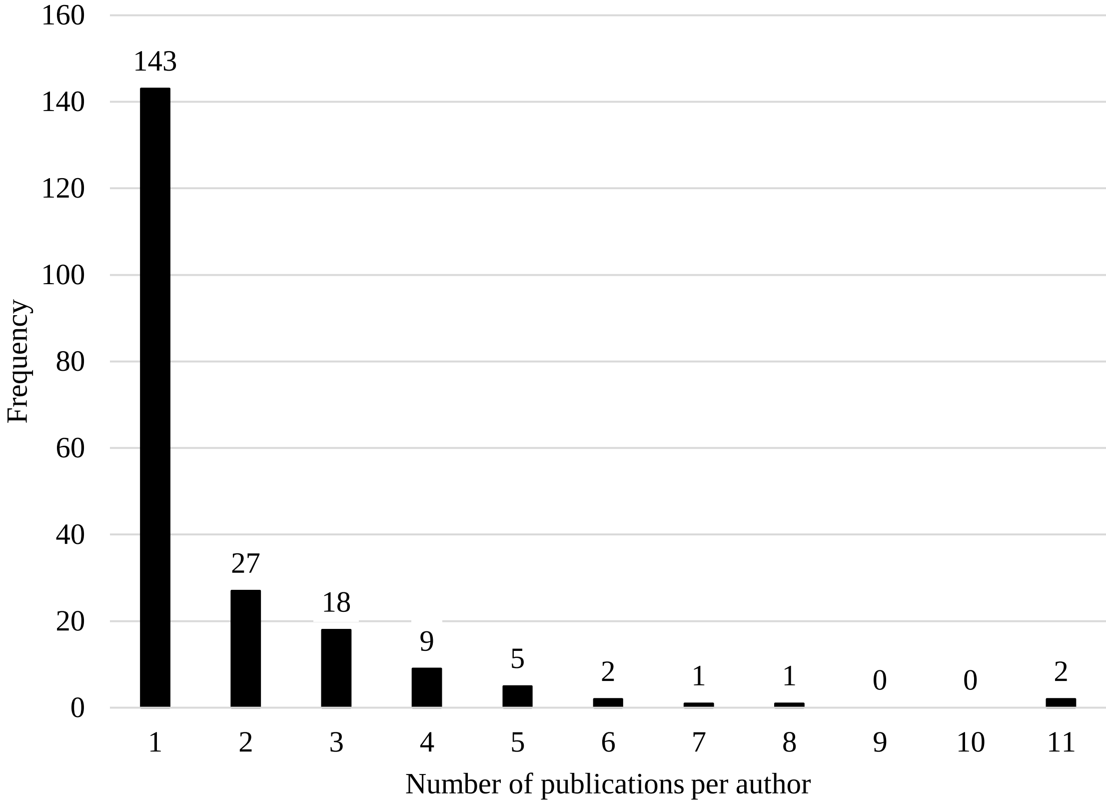
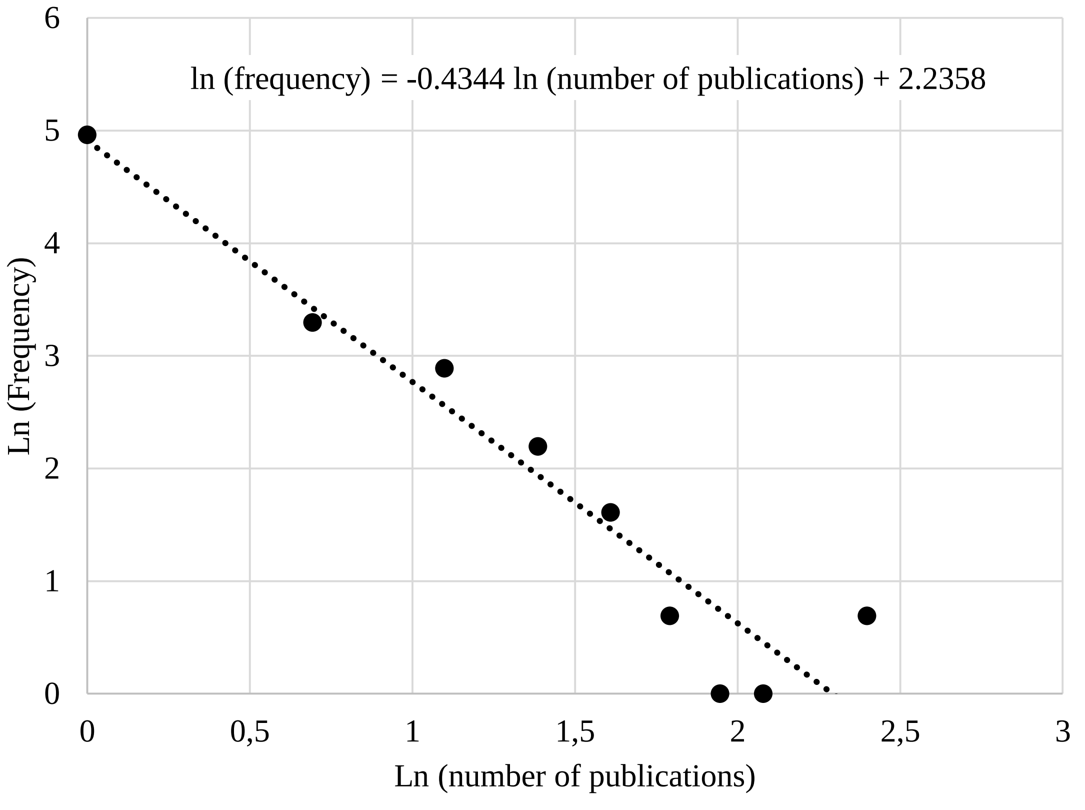
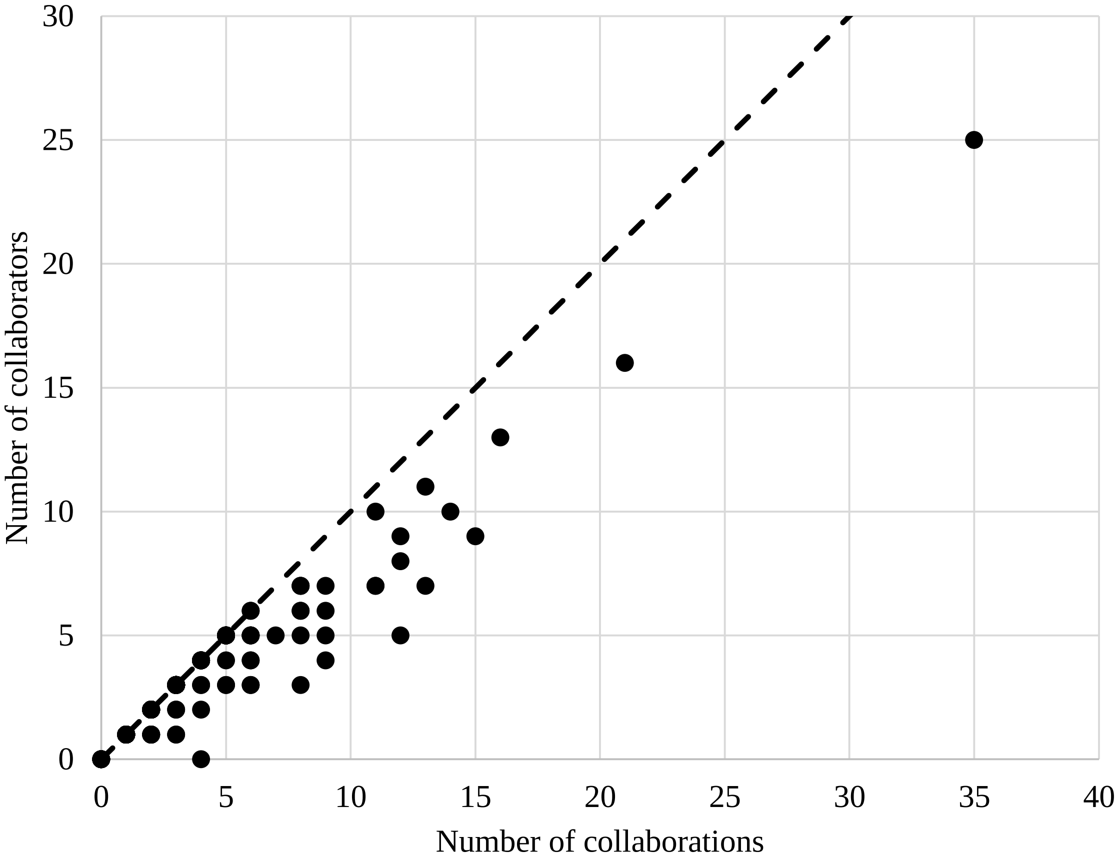


Figure 4.9: Power curve on the number of publications



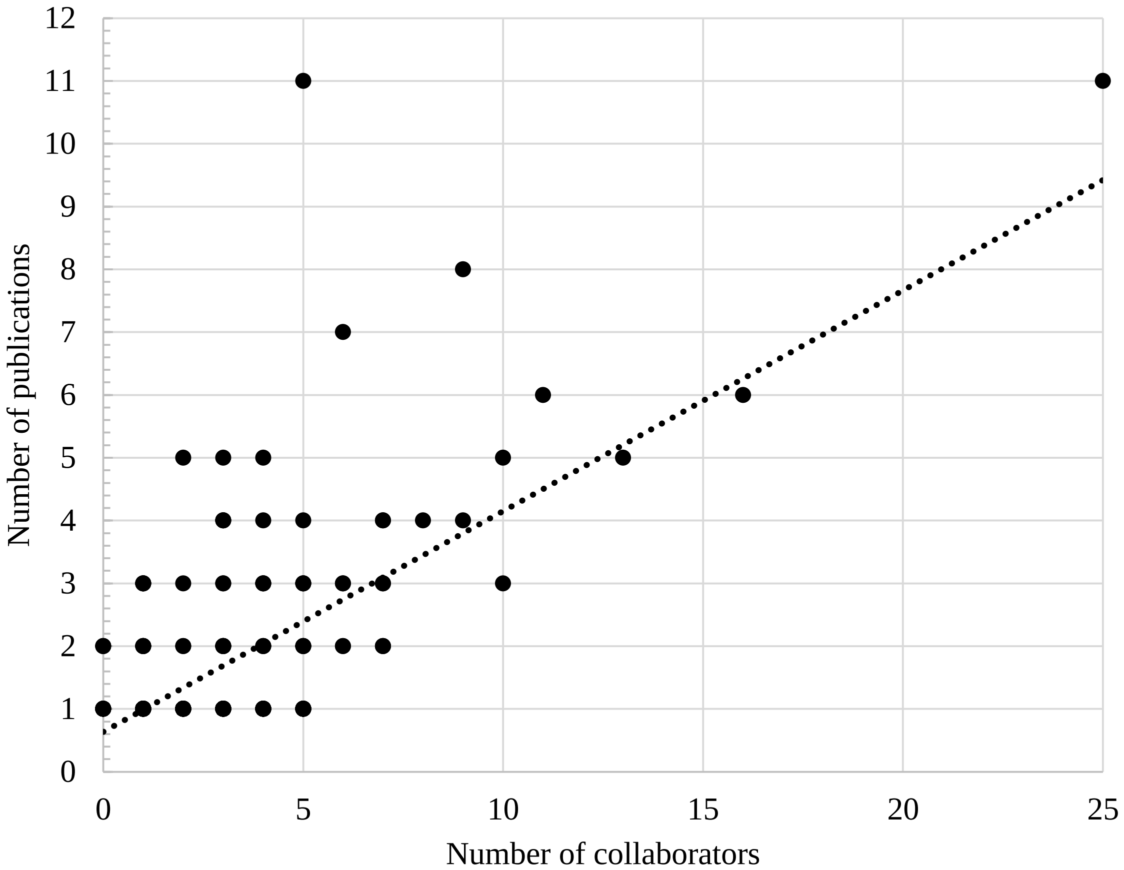
In order to evaluate the influence of research networks on the productivity of authors, two different measures were analyzed as a part of the systematic literature review, the number of collaborators and the number of collaborations (see Lee et al. 2014, p. 177). The former is a proxy for the broadness of an author's research network – it answers the question of how many different researchers this author has published with. To measure the intensity of joint research, the number of an author's collaborations can be used, which answers the question of how many (eventually also same) co-authors he had in total (see Lee et al. 2014). Figure 4.10 shows the values of the two numbers and their relationship for the data set. The most north-east point refers to L. Liang; he collaborated with 25 different researchers, and his intensity of joint research is captured by 35 collaborations.

Figure 4.10: Relationship between the number of collaborators and collaborations



In addition, Figure 4.11 compares the number of publications and the number of collaborators. In order to verify the hypothesis of Lee et al. (2014) that a wider collaboration network results in a higher number of publications (i.e., research output or productivity) the Spearman's rank correlation coefficient has been computed for the two variables. This non-parametric test allows determining the degree of interrelation between ordinal data sets.[64] In consideration of the results in Figure 4.11, the Spearman's rank correlation coefficient is 0.456, meaning that there is only a medium degree of correlation between the two variables. This is an interesting result as it is somehow counterintuitive and may be mainly justified with the aforementioned heterogeneity of the data set considered here.

Figure 4.11: Relationship between the number of publications and collaborators



In order to analyze the scientific influence of each article, a citation-based analysis was conducted. In a first step, the number of received citations was counted for all 135 DEA articles. Thereby, only citations from other hybrid or centralized approaches were taken into account. In a second step, the ranking of the most influential publications thus derived was compared to an article's number of citations at Google Scholar.[65]

---

64 For more details about the Spearman test, see e.g., Miah (2016).

65 The Google Scholar enquiries were conducted on the 10th of September 2018.

The results of both methodological steps are given in Table 4.6 for the 13 most cited publications.[66] Column 1 contains the authors and the year of the publication. Column 2 presents the respective number of citations from other hybrid or centralized DEA approaches. The last Column (i.e., Column 3) provides the corresponding quantity of citations at Google Scholar and, in addition, informs (in parenthesis) about the position of each article were the ranking to be derived according to this indicator instead.

Table 4.6: The publications with the most citations

	Number of citations …	
Publication	from other hybrid or centralized DEA approaches	at Google Scholar (ranking position)
1 Roll et al. (1991)	41	492 (1)
2 Roll and Golany (1993)	28	301 (3)
3 Kao and Hung (2005)	25	200 (7)
4 Beasley (2003)	24	277 (4)
5 Karsak and Ahiska (2005)	20	125 (12)
6 Liu and Peng (2008)	18	217 (5)
7 Jahanshahloo et al. (2005)	15	119 (13)
8 Sinuany-Stern and Friedman (1998)	14	196 (8)
9 Amin and Toloo (2007)	13	145 (10)
10 Athanassopoulos (1995)	12	145 (10)
11 Lozano and Villa (2004)	11	208 (6)
12 Cook and Kress (1990)	10	359 (2)
12 Friedman and Sinuany-Stern (1997)	10	174 (9)

One can conclude from Table 4.6 that the publications of Roll et al. (1991) and Roll and Golany (1993) received the most citations from other hybrid or centralized DEA approaches. Furthermore, these approaches would be placed at positions 1 and 3, respectively, when they are ranked according to the number of received citations at Google Scholar. However, Table 4.6 also indicates that the citation quantity at Google Scholar is not always a good indicator for the importance of a publication within a defined research area. For example, the article of Karsak and Ahiska (2005) only comprises 125 citations at Google Scholar which corresponds to position 13 in the

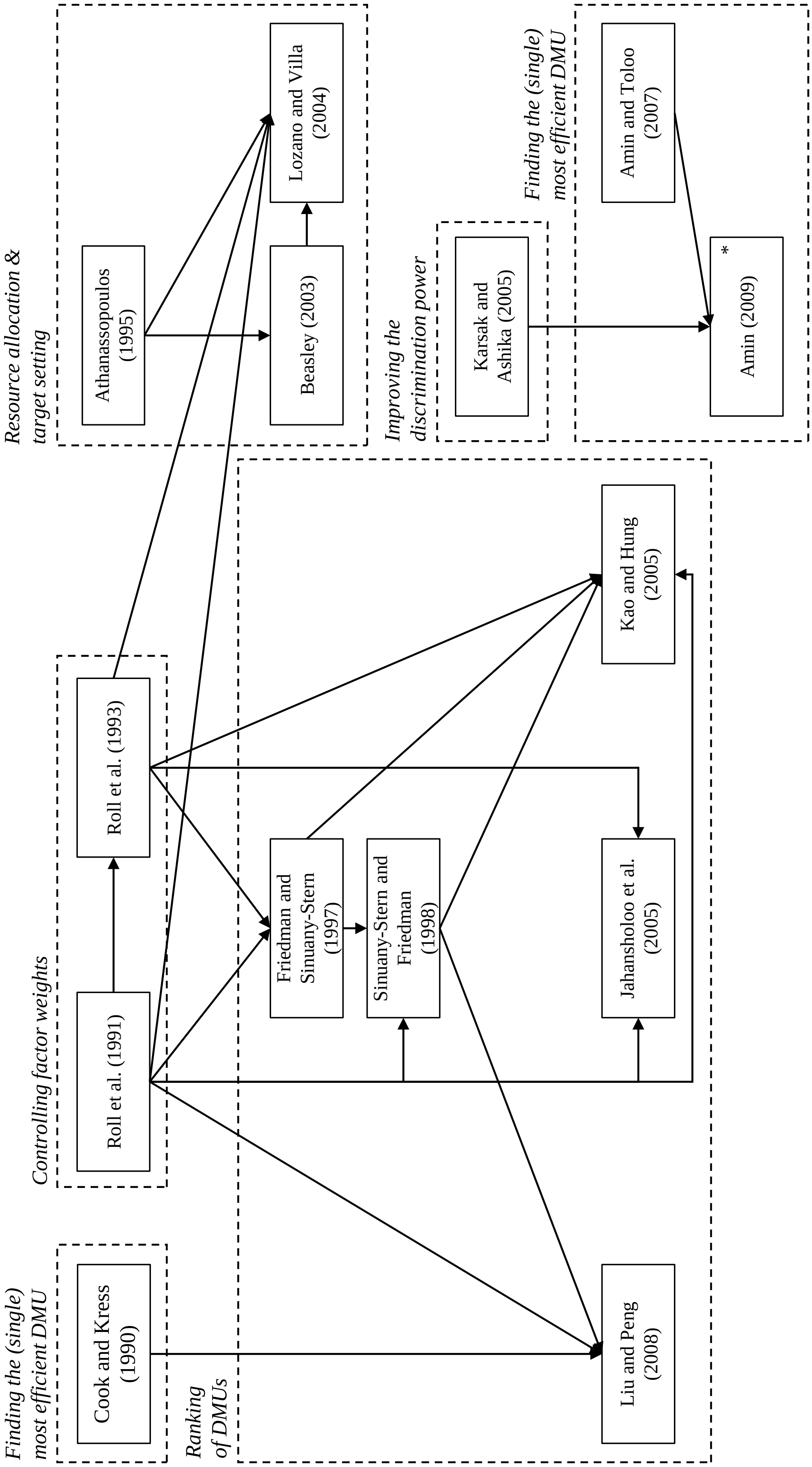
[66] Since numerous publications were placed on position 14 in the derived ranking, only the 13 most cited publications are discussed in more detail.

Google Scholar ranking. However, the article is placed on position 5 when only the relevant (i.e., hybrid and centralized) DEA articles are considered. The opposite is the case for the publication of Cook and Kress (1990): whereas this article is placed at position 12 according to the citations from other hybrid or centralized DEA approaches, it received the second most citations at Google Scholar. Interestingly, the Google Scholar ranking does not correspond to the year of publication either. For example, the article of Athanassopoulos (1995) has fewer citations at Google Scholar than the publication of Liu and Peng (2008) which has been published 13 years later.

The relationships between the most influential publications as provided by Table 4.6 are graphically illustrated in Figure 4.12. The arrows depict the reciprocal citation patterns and, therefore, can be straightforwardly interpreted as an article's influence on subsequent studies. Again, one can clearly see that the articles of Roll et al. (1991) and Roll and Golany (1993) have enormous influence on other research streams as the majority of the approaches considered here cite these two publications (see the depicted arrows in Figure 4.12). By contrast, only one article (e.g., Liu and Peng 2008) refers to the publication of Cook and Kress (1990), which is – chronologically – the first approach that has applied a common set of weights in a DEA.[67]

[67] Even other approaches classified under the same research stream (i.e., the contributions of Amin et al. 2006 as well as Amin and Toloo 2007) do not refer to the article of Cook and Kress (1990).

Figure 4.12: Citation-based relationships between the most influential papers



* The publication of Amin (2009) has not received enough references to be listed here. However, it is the corrigendum of Amin/Toloo (2007).

In order to evaluate the interdependencies between the different research streams more precisely, a further citation-based analysis was conducted: in a first step, the reciprocal references between the different hybrid and centralized DEA articles were traced. The results of this first step were documented in a “who is citing whom”-matrix with 135 columns and 135 rows (one column and row for each publication). In a second step, the matrix was aggregated according to the respective research stream of each publication. The resulting matrix has eight columns and eight rows (one column and row for each research stream). In a last step, the number of reciprocal references between the different research streams (as provided by the aggregated matrix) were visualized by means of a Sankey diagram.

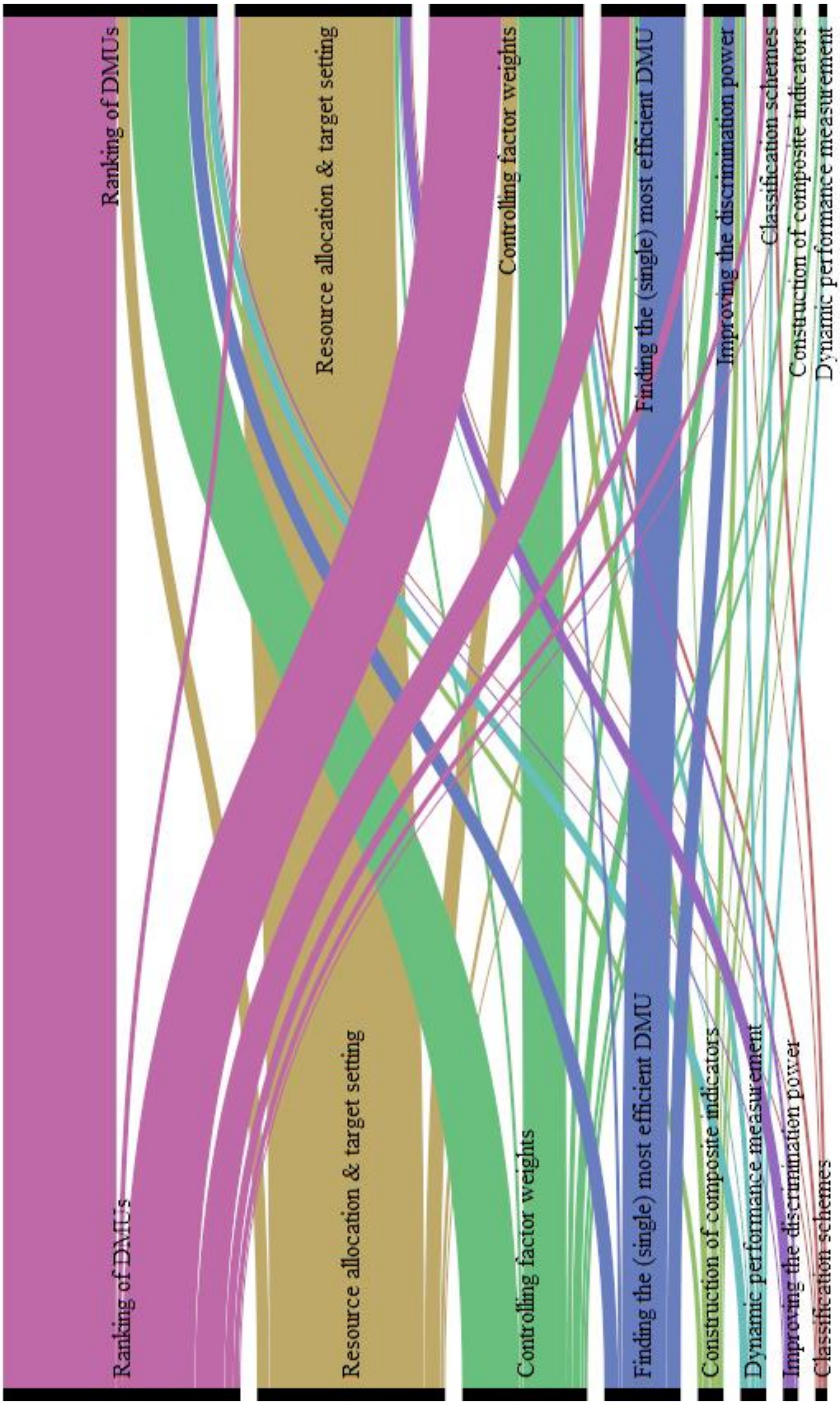
Sankey diagrams are usually applied in the area of engineering or logistics and use arrows to visualize the flow of material and energy quantities between different production units.[68] Thereby, the width of arrows is proportional to the represented material and energy quantities (see Schmidt 2008). In order to illustrate the results of the aforementioned citation-based analysis, the size of arrows was scaled according to the number of received citations. Hence, the arrows between the different research streams can be straightforwardly interpreted as the level of interdependency. The corresponding Sankey diagram is shown in Figure 3.11.

Interestingly, a substantial share of the citations of the research stream *Ranking of DMUs* stems from the category *Controlling factor weights*, and vice versa. Hence, it can be concluded that there exists an intense scientific relationship between these two application areas. By contrast, publications of the research stream *Resource allocation and target setting* receive the majority of references from publications associated with the same category. Hence, only a small share of these approaches refers to concepts that have already been applied in other application fields of DEA. This can be interpreted as an interesting research gap. However, it may also indicate the existence of particularities of the research stream *Resource allocation and target setting.*

---

[68] Examples of the application of Sankey diagrams in the field of engineering and logistics can be found, e.g., in Khurana et al. (2002).

Figure 4.13: Citation-based relationships between the research streams



4.5 Conclusions and research questions

Even though a consistent performance measurement framework needs to incorporate organizational structures to receive meaningful empirical results, a holistic literature review of how the perspective of organization theory is currently modeled in DEA is missing. The study presented here is designed as a first step in line with this research gap and intends to provide an overview of how different degrees of centralization are (implicitly or explicitly) incorporated in DEA.

To meet this research objective, three distinct levels of centralization were defined according to the attached input-output weights. The extreme case – a complete centralization – does not allow any freedom of decision and, hence, forces each DMU to operate according to the preferences of the central decision maker. In DEA, this is mathematically expressed through the uniform application of a common set of weights to all DMUs under consideration. The other extreme – complete decentralization – is implicitly assumed by basic DEA approaches. Their corresponding mathematical models allow each DMU to autonomously operate according to their individual preferences, which is expressed through the ability to choose their most favorable set of weights.

Being aware that in practical situations one can observe neither a complete centralization nor a complete decentralization of decisions, a compromise solution approach called "hybrid management scenario" was also defined. In hybrid management scenarios, the central decision maker allows flexibility to some extent regarding the DMU's decision making and simultaneously reserves the authority to take specific decisions at the hierarchical top level (e.g., strategic decisions). Hence, such organizations incorporate substantial characteristics of both management concepts. This can be implemented in DEA using combinations of complete weight flexibility and a common set of weights.

Based on a systematic literature review, 135 different approaches that (implicitly or explicitly) assume a hybrid or centralized management scenario were identified. According to the respective objectives of each article, the publications were categorized into eight distinct research streams. Due to the extensive amount of different approaches, only the most popular publications (in terms of citations from other hybrid or centralized DEA articles) were explained in detail.

The variety of hybrid as well as centralized DEA approaches and their contributions to eliminating shortcomings of conventional models provide clear evidence for their substantial relevance in the field of relative performance measurement. Despite the broad spectrum of these approaches, there will still be many research challenges to cope with. And although the other organizational variables stressed in Section 4.2 have not been connected to the DEA literature here, it can be assumed that their systematic appraisal will also reveal a wide range of further research opportunities.

The case of KONE Corporation provides a good example of how the specific organizational context raises yet unaddressed questions about how to appropriately measure the DMUs' efficiencies. The challenge was not so much to find topics but rather to set priorities. In consideration of the given organizational background and discussions with four representatives of the focal organization KONE, the following two research questions were elaborated. They are addressed in Chapters 5 and 6 below, respectively:

Research question 1: How can we measure the performance of KONE's maintenance units over time while taking into account the individual characteristics of each maintenance group represented by different group technologies over time?

> In many theoretical and practical contexts, the central decision maker will be interested in evaluating the productivity change of DMUs over time. To meet this objective, the literature on DEA already comprises a variety of different approaches (e.g., Färe et al. 1992a, Färe et al. 1994, Afsharian and Ahn 2014). Among them, the metafrontier Malmquist index of Pastor and Lovell (2005), which was further elaborated by Oh and Lee (2010), has received considerable attention. However, this traditional performance measurement approach is based on strong premises, which are usually not satisfied in practice. Among other things, the approach assumes that combinations of local technologies (as represented by the convex metafrontier) are producible and, hence, attainable for each operating entity. This is clearly inconsistent when the organization seeks to implement distinct and customized strategies and, therefore, makes extensive use of the concept of specialization. In the exemplary case of KONE's German subsidiary, the overall maintenance tasks are allocated to different regions (i.e., region-

oriented specialization) to receive efficiency gains from learning effects and focused customer strategies (see Section 2.1). However, the metafrontier Malmquist index would implicitly neglect this organizational issue. Consequently, this approach may lead to an incorrect approximation of the metafrontier and accordingly to misleading results and managerial conclusions. Improving the estimation of the metafrontier, an alternative approach is proposed that preserves the individual characteristics of each technology. This unique feature of the proposed approach makes it applicable to situations where the organization makes extensive use of the concept of specialization and, hence, provides valuable managerial outcomes for further analyzing productivity changes over time.

Research question 2: How can we compare the performance of KONE's management groups while maintaining the individual characteristics of each maintenance group represented by different group technologies?

Besides measuring productivity changes of individual DMUs in a dynamic setting, a central decision maker is usually also interested in identifying performance differences between groups of DMUs. Such comparisons can yield valuable managerial information, e.g., about superior management styles or customized strategies.

So far, the DEA-literature does not provide a method that is able to compare groups of units in centralized management scenarios. However, for decentralized management scenarios, Camanho and Dyson (2006) proposed a Malmquist index-based approach, which has already been introduced in the course of Section 3.4.3. Based on their framework for comparing the performance of groups, a corresponding approach is proposed for centralized management environments. The framework introduced allows straightforwardly comparing KONE's distinct management groups while preserving the individual characteristics of each group technology. Hence, this approach can also yield important information regarding the productivity of different strategies.

5 A non-convex metafrontier Malmquist index for measuring performance changes over time[69]

5.1 Introduction

Since the introduction of the Malmquist index by Färe et al. (1992a) (see Section 3.4), a few limitations have been faced by researchers. As this form of the Malmquist index uses the geometric mean of two measures of productivity change – which refer to the adjacent time periods under consideration – it fails circularity. Infeasibilities can also occur when DEA models under VRS are used to compute and decompose the index. Over the last two decades, the depicted shortcomings have motivated researchers to focus on the methodological development of the Malmquist index and its decomposition. A thorough review of the family of the Malmquist indices can be found in Afsharian and Ahn (2015).

Among the different frameworks of the Malmquist index, the metafrontier Malmquist index has recently begun to receive considerable attention by researchers. The reason is that not only this form of the Malmquist index can overcome the above-mentioned issues but also offers a number of other interesting features. This index was proposed first by Hayami and Ruttan (1970) and developed further in the area of SFA for the estimation of technical efficiencies and technology gaps for observations that may not have the same technology. Their approach assumes that – within the same industry – there are several well-defined groups of observations, which operate under their own local technologies. Accordingly, local frontiers are constructed by considering all observations belonging to

[69] A slightly modified version of this chapter has been published as Afsharian, M., H. Ahn, S. G. Harms. 2018a. A non-convex meta-frontier Malmquist index for measuring productivity over time. *IMA Journal of Management Mathematics*. Vol. 29(4), pp. 377-392.

the same group while the metafrontier is the envelope of the group frontiers. The primal version of this index no longer measures the productivity change between two time periods, but provides a cross-sectional comparison of the performance of groups of DMUs in a static setting. Therefore, it has recently been enhanced by Pastor and Lovell (2005) and Oh and Lee (2010) as a tool to also measure productivity change over time (see also Portela and Thanassoulis 2008, Oh 2010, Afsharian and Ahn 2015).

According to the design of the metafrontier Malmquist index, a single metatechnology is constructed from data for observations belonging to all groups and observed in all time periods. This metatechnology then serves as a “global” benchmark, representing the best experienced technology among all groups and over all time periods in the analysis. On this basis, one not only can measure the within-group efficiency of units at a specific period of time but also capture how their efficiency has changed with regard to the metatechnology. Although the central concept of this index is compelling, it suffers from a drawback: the metatechnology is formed by the convex union of all experienced group technologies over time. Taking into account even a static setting (where only a cross-sectional analysis is applied), researchers argue that any metatechnology, which is formed as the union of “even convex group technologies”, is unlikely to be convex (see e.g., Huang et al. 2013, Kerstens et al. 2019). This can be more problematic when the index also includes a time component to measure productivity change over a number of time periods. In this case, the convexification neglects that the technology under which each group of units operates can change over time. This negligence may lead to an incorrect estimation of the metafrontier such that the corresponding results of productivity will not properly reflect the performance.

Against this background, this chapter introduces a new way of estimating the metatechnology, which applies the minimum extrapolation principle on the aggregation of the experienced group technologies over time. As will be shown, the resulting index, called the non-convex metafrontier Malmquist index, provides more accurate results compared to the existing metafrontier Malmquist index. The new index also preserves the role of each group technology – observed at the specific time period – in the estimation of the metatechnology. Therefore, individual characteristics of the group technologies can later be traced in measuring productivity change. In particular, this includes information about group technologies which contribute significantly to the shape of the metatechnology over time. This unique feature of the suggested approach plays a crucial role in measuring and analyzing productivity, where a further diagnosis of individual performances is required. With respect to both computational and test properties, the proposed index also possesses the circularity property, generates a single measure of productivity change and is immune to infeasibility under VRS. Similar to traditional indices, it can be decomposed into the standard components such as efficiency change and best practice change.

The rest of this chapter unfolds as follows: after some preliminaries and technical background in Section 5.2, the idea of estimating the metafrontier technology is depictured in Section 5.3. In Section 5.4, the proposed metafrontier Malmquist productivity index is formulated mathematically. Section 5.5 illustrates the new index and its properties by means of an empirical application to KONE Corporation. The chapter concludes with a summary and an outlook on future research opportunities in Section 5.6.

5.2 Technical background

Suppose that there exists a panel of *n* DMUs which can be partitioned into *G* (*G*>1) distinct groups observed in *T* time periods. Let each group *g* (*g* = 1,…,*G*) include δ_g DMUs $(X_j^{g,t}, Y_j^{g,t}) \in \Re_+^m \in \Re_+^s$ ($j = 1, ..., \delta_g$), where $X_j^{g,t} = (x_{1j}^{g,t}, x_{2j}^{g,t}, ..., x_{mj}^{g,t})$ and $Y_j^{g,t} = (y_{1j}^{g,t}, y_{2j}^{g,t}, ..., y_{sj}^{g,t})$ are non-negative and non-zero vectors of inputs and outputs, respectively, observed in period *t* (*t* = 1,…,*T*). Following O'Donnell et al. (2008), it is assumed that all DMUs in each group *g* operate under the same technology, resulting from, e.g., the same resource, regulatory or other environmental constraints. Hence, each local technology of group *g* in time period *t* can be represented by a PPS (in the following also abbreviated as "technology") of feasible input-output combinations as follows:

$$PPS^{g,t} = \left\{ (X^{g,t}, Y^{g,t}) \in \Re_+^m \times \Re_+^s \,\middle|\, X^{g,t} \ can\ produce\ Y^{g,t} \right\}. \tag{5.1}$$

Throughout this chapter, without loss of generality, it is assumed that the local technologies in (5.1) satisfy non-emptiness, free disposability, convexity and minimum extrapolation. The following analysis may be straightforwardly extended to other types of technologies with other axioms. Taking into account these axioms, the local technologies in (5.1) can be expressed precisely by means of the following technology sets:

$$PPS^{g,t} = \left\{ \begin{array}{l} (X^{g,t}, Y^{g,t}) \in \Re_+^m \times \Re_+^s \mid x_i^{g,t} \geq \sum_{j=1}^{\delta_g} \lambda_j^{g,t} x_{ij}^{g,t},\ y_r^{g,t} \leq \sum_{j=1}^{\delta_g} \lambda_j^{g,t} y_{rj}^{g,t}, \\ \sum_{j=1}^{\delta_g} \lambda_j^{g,t} = 1, \quad \lambda_j^{g,t} \geq 0, \quad j = 1,\ldots,\delta_g \end{array} \right\}. \tag{5.2}$$

With respect to the definition of $PPS^{g,t}$ in (5.2), one can measure the efficiency of a DMU against the frontier of a particular group g (g = 1,…,G) at the specific point of time t (t = 1,…,T). Moreover, one may also measure a "within-period" efficiency by means of a contemporaneous benchmark technology as (see e.g., O'Donnell et al. 2008, Huang et al. 2013)

$$PPS^{M,t} = PPS^{1,t} \cup PPS^{2,t} \cup,\ldots,\cup PPS^{G,t} \tag{5.3}$$

where $PPS^{M,t}$ is formed by the aggregation of all group technologies in time period t (t = 1,…,T). It should be noted that using (5.3) as a benchmark in a specific period t provides a cross-sectional comparison of the performance of groups of DMUs in a static setting, i.e., the measurement is done in a specific time period t. In order to make the comparison dynamic, (5.3) has to be modified. This can be done by the concept of the metafrontier Malmquist index as follows:

Based on, e.g., Oh and Lee (2010), the metafrontier Malmquist index for a DMU p which belongs to group g (g = 1,…,G), regarding two time periods t and t+1, is defined as:

$$MI^M(X_p^{g,t+1}, Y_p^{g,t+1}, X_p^{g,t}, Y_p^{g,t}) = \frac{Eff^M(X_p^{g,t+1}, Y_p^{g,t+1})}{Eff^M(X_p^{g,t}, Y_p^{g,t})}. \tag{5.4}$$

In this formula, $Eff^M(X_p^{g,t}, Y_p^{g,t})$ and $Eff^M(X_p^{g,t+1}, Y_p^{g,t+1})$ represent the two required input-oriented meta-efficiencies, which can be determined as:

$$Eff^M(X_p^{g,k}, Y_p^{g,k}) = \min\left\{ \theta_p^k : (\theta_p^k X_p^{g,k}, Y_p^{g,k}) \in PPS^M \right\}, \quad k = t, t+1. \tag{5.5}$$

PPS^{M} is the metatechnology, which aggregates all group technologies over all time periods as

$$PPS^{M} = \bigcup_{t=1}^{T} \bigcup_{g=1}^{G} PPS^{g,t}. \tag{5.6}$$

More details about the metafrontier Malmquist index and its potentials to provide interesting insights into DEA applications can be found in a series of papers such as in Portela and Thanassoulis (2008), Oh (2010), Chen and Yang (2011), Portela et al. (2011), Afsharian and Ahn (2015), Choi et al. (2015) as well as Kerstens et al. (2019).

5.3 Motivation and graphical explanations

In the existing form of the metafrontier Malmquist index, all observations from all groups in all periods are assumed to be theoretically and potentially able to access a single best practice technology. This metatechnology – which is assumed to be available to the whole industry in which the DMUs operate – is then obtained by the "convex aggregation" of the group technologies over time (see e.g., Oh and Lee 2010, Chen and Yang 2011). On this basis, all observations from different groups and time periods are accepted to form the meta-benchmark technology. This means that the characteristics of the group technologies are implicitly assumed to remain unaltered over time, i.e., it is assumed to be no technical differences between different groups of DMUs which are observed over time. This is clearly inconsistent with the primary setting of the problem by which the DMUs are partitioned to G distinct groups observed in T time periods. As a consequence, although observations in each time period can be considered to be acceptable to form the respective group technology set in a specific time period, including all observations from all groups in all periods in the analysis (to estimate the metatechnology) is questionable, as illustrated also graphically in the following.

Let us suppose that there exist two group technologies g and g+1 in a single time period t. It is also assumed that there are two inputs and a single output, and the observations have the same level of output. The corresponding local technologies $T^{g,t}$ and $T^{g+1,t}$

are depicted in Figures 5.1 and 5.2. According to the definition in (5.3), one can form a respective contemporaneous benchmark technology to provide a cross-sectional comparison of the performance of these two groups of DMUs in the single time period *t.* Following the existing method of aggregation, the resulting contemporaneous technology – indicated by $Convex\ T^{M,t}$ – will be the one shown in Figure 5.1. Considering the frontier of this technology set, one can see that there are areas – shown by "infeasible input combinations" – which have neither been experienced nor producible in practice. In fact, this area is only formed as a direct consequence of the imposed convexity assumption between these groups to easily estimate the contemporaneous technology $Convex\ T^{M,t}$.

Figure 5.1: Convex estimation of a contemporaneous technology set

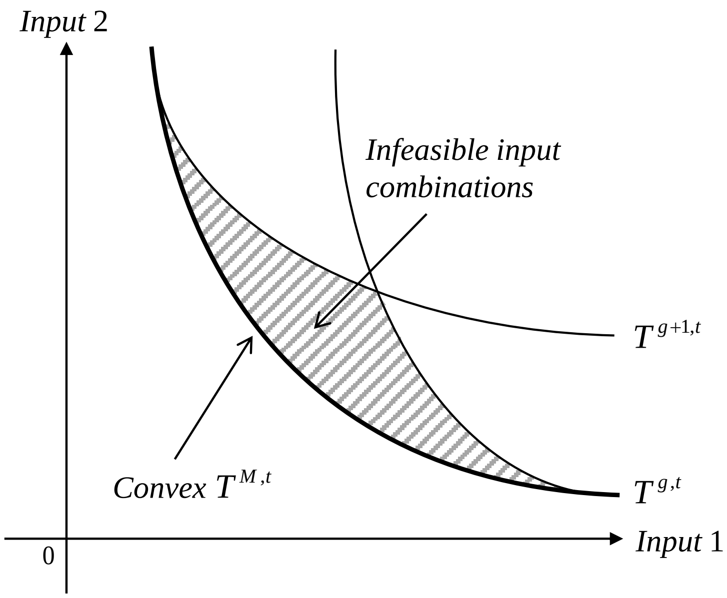
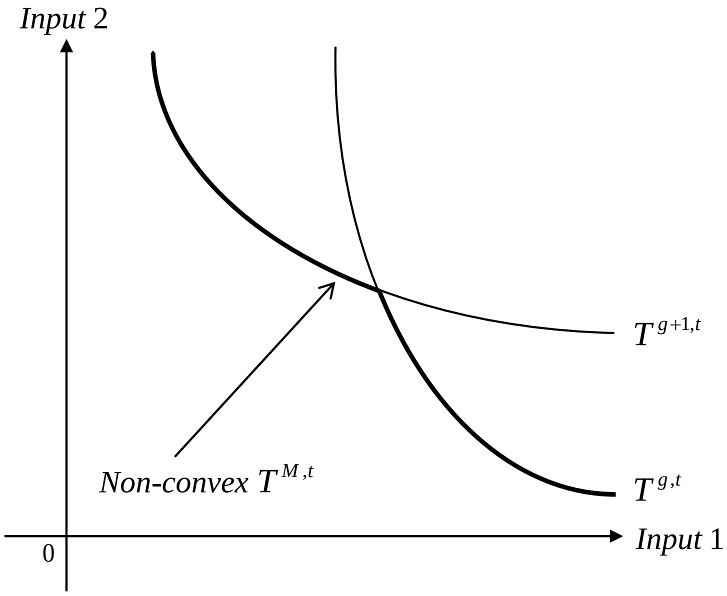


Figure 5.2: Non-convex estimations of a contemporaneous technology set



One can now compare this result to the more precise aggregation of the group technologies as shown in Figure 5.2 by *Non-convex* $T^{M,t}$. As can be seen here, the aggregation of the distinct group technologies $T^{g,t}$ and $T^{g+1,t}$ applies the minimum extrapolation principle. Hence, the resulting *Non-convex* $T^{M,t}$ provides a pure union of what has really been occurred rather than also having an additional area resulting from the convexification between these groups.

The above graphical example shows that assuming convexity even between observations originating from different group technologies is a strong premise, while in a multi-period analysis this phenomenon becomes more problematic. The reason is that not only the business environment but also, e.g., government rules or regulations, policy directives and economic conditions, under which the DMUs operate, can change significantly over time. Therefore, convex combinations of units belonging to different time periods may also once more reduce the accuracy of the estimation of the metatechnology. A graphical example of this can be seen in Figure 5.3, which depicts a metatechnology formed from the two group technologies *g* and *g*+1 over two time periods *t* and *t*+1.

As can be seen in Figure 5.3, the existing metafrontier Malmquist index proposes a metatechnology which is the convex envelope of all group technologies over time, i.e., the convex aggregation of $T^{g,t}$, $T^{g+1,t}$, $T^{g,t+1}$ and $T^{g+1,t+1}$, indicated by *Convex* T^{M}. This result can be compared to proposed estimation of the metatechnology *Non-convex* T^{M}

shown in Figure 5.4. A comparison between these two forms highlights how the convexification can produce virtual points which are the result of a poor estimation of the metatechnology, marked by the shaded area in Figure 5.3. It should be noted that even if contemporaneous technology sets (i.e., $T^{C,t}$ and $T^{C,t+1}$) were to satisfy convexity (e.g., perhaps in order to have a simple approximation), as one can expect that the environment can change over time, there would be no reason why the union of these contemporaneous technology sets should be convex to estimate the metatechnology. Hence, it can be concluded that the proposed *Non-convex* T^M as the metatechnology is a more accurate and appropriate estimate of the best practice technology, which has really been experienced over time.

In addition to the above remark about the estimation of the metatechnology, a closer look at the structure of *Convex* T^M and *Non-convex* T^M in Figures 5.3 and 5.4 reveal another unique feature of the new form of the benchmark technology. A comparison between both forms shows that the proposed benchmark technology preserves the role of each group technology – observed at a specific time period – in the estimation of the metatechnology, i.e., information about local group technologies are not mixed. Unlike in *Convex* T^M, the proposed approach allows tracing individual characteristics of the group technologies while measuring productivity change. In particular, information about group technologies, which contribute significantly to the shape of the metatechnology (in the following called “superior group technologies”), is revealed. This property of the *Non-convex* T^M plays a crucial role in measuring and analyzing productivity, where a further diagnosis of individual performances is required.

Figure 5.3: Convex estimation of a metatechnology set

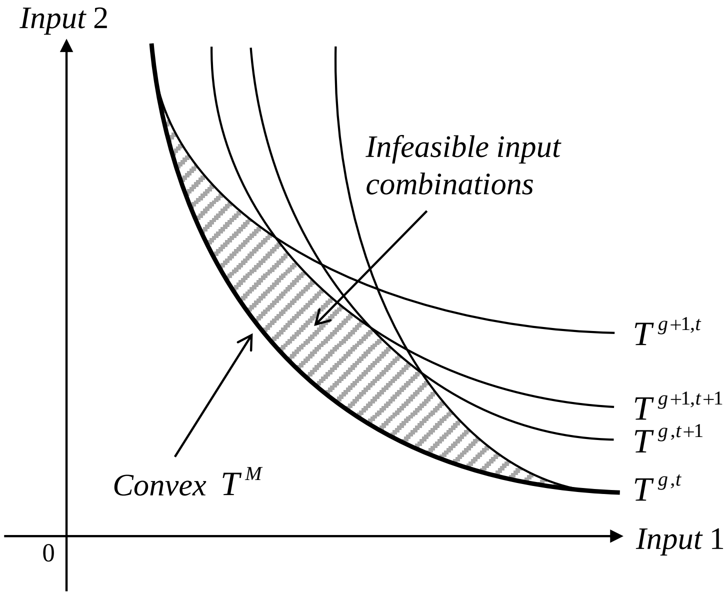
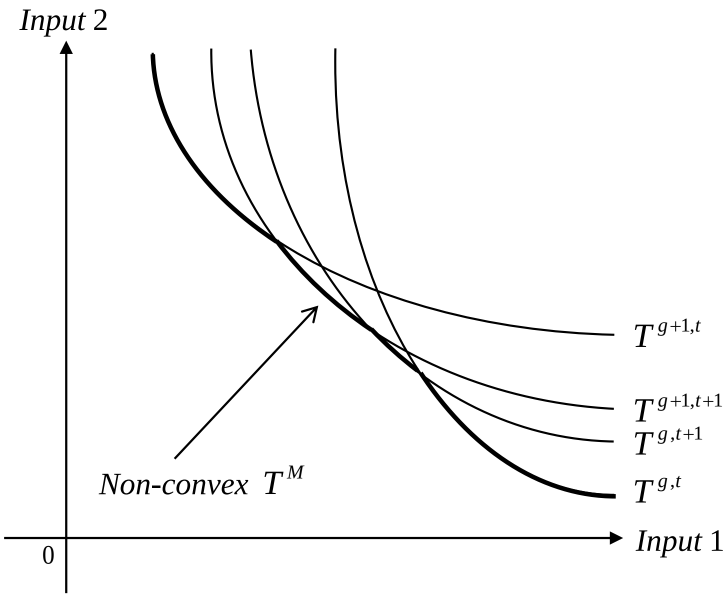


Figure 5.4: Non-convex estimation of a metatechnology set



5.4 The proposed approach

In this section, the metafrontier Malmquist productivity index based on the proposed metatechnology is introduced. In accordance with the graphical example given in Section 5.3, this metatechnology is formed by the pure union of all local technologies in (5.3) as

$$PPS_{NC}^{M} = \bigcup_{t=1}^{T}\bigcup_{g=1}^{G} PPS^{g,t} \tag{5.7}$$

where PPS_{NC}^{M} has been denoted by a subscript "NC" to emphasize that the metatechnology is now formed based on the non-convex union of the group technologies. This representation of the metatechnology can be precisely modeled by considering a number of mathematical axioms as follows:

1. (Non-emptiness). The observed $(X_j^{g,t}, Y_j^{g,t}) \in PPS_{NC}^{M}$, g = 1,…,G; t = 1,…,T; j=1,…, δ_g .
2. (Free disposability). If $(X,Y) \in PPS_{NC}^{M}$, $X' \geq X$, $Y' \leq Y$, then $(X',Y') \in PPS_{NC}^{M}$.
3. (Local convexity). If (X,Y) and $(\tilde{X},\tilde{Y}) \in PPS_{NC}^{M}$, then $\lambda(X,Y) + (1-\lambda)(\tilde{X},\tilde{Y}) \in PPS_{NC}^{M}$ for any $\lambda \in [0,1]$, provided that there exists t (t = 1,…,T) and g (g = 1,…,G) such that both (X,Y) and $(\tilde{X},\tilde{Y}) \in PPS^{g,t}$.
4. (Minimum extrapolation). PPS_{NC}^{M} is the smallest set which satisfies axioms 1-3.

With regard to the standard assumptions of DEA models, the meaning of axioms #1 and #2 is obvious. According to axiom #3, convex combinations among members of different group technologies are not required. Axiom #4 then ensures that PPS_{NC}^{M} will be the smallest set, which results from the pure union of the local group technologies. Taking into account these axioms, the definition of the metatechnology in (5.7) can mathematically be enhanced as:

$$PPS_{NC}^{M} = \bigcup_{t=1}^{T}\bigcup_{g=1}^{G} \left\{ \begin{array}{l} (X^{g,t}, Y^{g,t}) \in \Re_{+}^{m} \times \Re_{+}^{s} \mid \, x_i^{g,t} \geq \sum_{j=1}^{\delta_g} \lambda_j^{g,t} x_{ij}^{g,t}, \; y_r^{g,t} \leq \sum_{j=1}^{\delta_g} \lambda_j^{g,t} y_{rj}^{g,t}, \\ \sum_{j=1}^{\delta_g} \lambda_j^{g,t} = 1, \quad \lambda_j^{g,t} \geq 0, \quad j = 1,\ldots,\delta_g \end{array} \right\}. \tag{5.8}$$

On this basis, $Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t})$, which captures the input-oriented efficiency of unit p belonging to a group g with the data from period t, can be determined as:

$$Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t}) = \min\left\{ \theta_{p(g,t)}^{M} : \left(\theta_{p(g,t)}^{M} X_p^{g,t}, Y_p^{g,t}\right) \in PPS_{NC}^{M} \right\}. \tag{5.9}$$

In order to compute the meta-efficiencies in (5.9), let us first assume that $Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t})$ (q = 1,…,G and l = 1,…,T) represent all local efficiencies of unit p (belonging to a group g with the data from period t) with the following definition:

$$Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t}) = \min\left\{\ \theta_{p(g,t)}^{q,l} : \left(\theta_{p(g,t)}^{q,l} X_p^{\,g,t}, Y_p^{\,g,t}\right) \in PPS^{\,q,l}\ \right\}, \qquad q = 1,\dots,G;\ \ l = 1,\dots,T. \tag{5.10}$$

According to (5.10), one can measure these local efficiencies of a unit p against any group technology q in time period l with the following linear programming problems:

$$Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t}) = min\left\{\theta_{p(g,t)}^{q,l}\ \middle|\ \begin{array}{ll} \sum_{j=1}^{\delta_g} \lambda_j^{g,t} x_{ij}^{g,t} \le x_{ip}^{q,l}\, \theta_{p(g,t)}^{q,l}, & i = 1,\dots,m \\ \sum_{j=1}^{\delta_g} \lambda_j^{g,t} y_{rj}^{g,t} \ge y_{rp}^{q,l}, & r = 1,\dots,s \\ \sum_{j=1}^{\delta_g} \lambda_j^{g,t} = 1 & \\ \lambda_j^{g,t} \ge 0, & \end{array}\right\}. \tag{5.11}$$

Now with respect to the discrete nature of the metatechnology in (5.8), the meta-efficiencies in (5.9) can be computed by the following enumeration procedure:

$$Eff_{NC}^{\,M}(X_p^{\,g,t}, Y_p^{\,g,t}) = \min_{\substack{l=1,\dots,T \\ q=1,\dots,G}} \left\{ Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t}) \right\}. \tag{5.12}$$

In this procedure, determining $Eff_{NC}^{\,M}(X_p^{\,g,t}, Y_p^{\,g,t})$ against the metatechnology is identical with finding the minimum value among $Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t})$ for all l (l = 1,…,T) and all q (q = 1,…,G) in which $Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t})$ can also be computed in advance by the corresponding DEA models in (5.11). It should be noted that as the DMU under evaluation is a real unit, at least one of its within-group efficiencies $Eff^{\,q,l}(X_p^{\,g,t}, Y_p^{\,g,t})$ is feasible, e.g., it is enveloped by the technology in which it has been observed. According to (5.8), this guarantees that $Eff_{NC}^{\,M}(X_p^{\,g,t}, Y_p^{\,g,t})$ is feasible. However, the above proposed formula (5.12) enumerates in its procedure all within-group efficiencies including those which might be infeasible for this unit. This can occur when DMU p is not enveloped by the boundary of a particular group technology at a specific period of time. For overcoming this problem in the computation of (5.12), such infeasible results of efficiency can be replaced in advance by sufficiently big values.

According to the graphical examples in Section 5.3 and also the way the metatechnology is formed by the convex and non-convex approaches, one can formulate $\textit{Non-convex}\ T^{M} \subseteq \textit{Convex}\ T^{M}$. On this basis, $Eff^{M}(X_p^{g,t}, Y_p^{g,t}) \leq Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t})$ where $Eff^{M}(X_p^{g,t}, Y_p^{g,t})$ and $Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t})$ denote the meta-efficiency of the convex and the non-convex approaches, respectively. This shows that a sufficient condition for the equality of these two approaches is the convexity of the metatechnology. Therefore, – as an extreme case from a theoretical point of view – if the grouping of units over time leads to a convex shape of the metatechnology, the efficiency results of these approaches will be exactly the same. However, in practical situations, the results tend to diverge when the metatechnology exhibits areas violating convexity in its shape.

Taking into account the definition of the best practice technology in (5.8), the proposed non-convex metafrontier Malmquist index for DMU p which belongs to group g, regarding two time periods t and t+1, is defined as

$$MI_{NC}^{M}(X_p^{g,t+1}, Y_p^{g,t+1}, X_p^{g,t}, Y_p^{g,t}) = \frac{Eff_{NC}^{M}(X_p^{g,t+1}, Y_p^{g,t+1})}{Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t})} \tag{5.13}$$

where $Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t})$ and $Eff_{NC}^{M}(X_p^{g,t+1}, Y_p^{g,t+1})$ represent the meta-efficiencies which can be computed by the formula in (5.12). The proposed metafrontier Malmquist index can also be represented by means of the following standard decomposition:

$$MI_{NC}^{M}(X_p^{g,t+1}, Y_p^{g,t+1}, X_p^{g,t}, Y_p^{g,t}) =$$
$$\underbrace{\frac{Eff^{g,t+1}(X_p^{g,t+1}, Y_p^{g,t+1})}{Eff^{g,t}(X_p^{g,t}, Y_p^{g,t})}}_{Efficiency\ Change} \times \underbrace{\left[\frac{Eff^{g,t}(X_p^{g,t}, Y_p^{g,t})}{Eff_{NC}^{M}(X_p^{g,t}, Y_p^{g,t})} \times \frac{Eff_{NC}^{M}(X_p^{g,t+1}, Y_p^{g,t+1})}{Eff^{g,t+1}(X_p^{g,t+1}, Y_p^{g,t+1})}\right]}_{Best\ Practice\ Change}. \tag{5.14}$$

The first component in (5.14) is the EC component. It captures the change in the technical efficiency of the unit under evaluation between two time periods t and t+1. The second component is the Best Practice Change (BPC) component, which indicates whether group technology g in time period t+1 in the region this unit operates is closer to or farther away from the metafrontier than is group technology g in time period t. On this basis, if the value of the metafrontier Malmquist index or any of its components is less than one, it denotes regress; a value greater than one implies progress, while a value of one indicates unchanged situation. Since the metatechnology is obtained by the aggregation of the group technologies – similar to the original version of the metafrontier Malmquist index –, the proposed index and its components are circular and immune to infeasibility under VRS.

5.5 Empirical illustration

5.5.1 Data set and model specification

In order to illustrate the proposed Malmquist index, the maintenance units of the company KONE are analyzed over the time period 2014-2015 (i.e., t = 2014, 2015). Due to data irregularities there were only 41 comparable maintenance units in the two respective time periods (i.e., n = 41). As it has been mentioned before, KONE has partitioned the maintenance units into four distinct managerial groups with regional headquarters in Hamburg, Berlin, Cologne and Munich, respectively.[70] For the sake of data anonymization, a randomly-selected number from 1-4 is given to each of these groups (i.e., G1, G2, G3 and G4).

In the context of this study, the representatives of KONE decided that two inputs and two outputs should be used to evaluate the maintenance units' operational efficiency. The inputs are *the number of full-time equivalent employees (FTE)* and *weighted response time (WRT).* The first measures the number of employees in respect of total hours worked,

[70] See Section 2.1 for more information on KONE's maintenance units and managerial groups.

whereby a FTE of 1.0 equals a full-time employee. The WRT adds up the time needed for maintenance tasks, repair tasks and the elimination of unexpected malfunctions with the relative weights of 0.2, 0.2 and 0.6, respectively. The two outputs are *the number of callouts (NOC)* and *total handling tasks (THT).* NOC is an indicator to measure the work quality of the maintenance units. It represents the number of registered customer orders and complains due to any interruptions of lift machines or a repair service with poor quality. Since both the customers and the company itself expect a minimum number of callouts, the NOC is considered as an undesirable output, which has to be minimized. The indicator THT, as the second output, presents the total number of installations to be maintained or repaired by the respective maintenance units. Since the company seeks to increase the number of commissioned tasks per unit (without further deteriorating the level of the other variables), this indicator is taken into account as an output to be maximized.

With respect to the theory put forward in Section 5.3, assuming convexity between observations within four group technologies is a strong premise even when a cross-sectional analysis is done, while in a multi-period analysis, as in the case of KONE, this phenomenon becomes more problematic. In such a dynamic setting, the internal and external business environment, under which the four maintenance groups operate, can be expected to change over time.

A noteworthy example is the change in personnel positions in the top and regional management of KONE, which occurred during the course of 2014 (see KONE 2017a). Such changes often have a considerable effect on the strategy and policy directives of the whole company. It is therefore questionable if the managerial approaches used in 2014 are identical with those applied in the subsequent year. Moreover, high market dynamics over the period under study has caused a lift inventory growth of 2.64 %. Since the data set does not show any substantial changes in terms of the employed personnel, it can be deduced that the unit managers would have restructured their processes (e.g., optimization of the route planning) to handle their additional work load. Furthermore, a new law for the operation of lifts (i.e., the so-called industrial safety regulation) came into force in June 2015. The new law provided, among others, more frequent audits and also prescribed more restricted regulations regarding the maintenance of lifts (see KONE 2017b). This should have demanded more time to invest and thorough work of the technicians at the time.

In addition, one should also take into account some other changes in the external environment, such as changes in government rules or regulations, policy directives and economic conditions over the time period 2014-2015. These changes are also likely to make combinations of maintenance units from different time-periods unreasonable. As a consequence, including all convex combinations of all observations in all time periods in the analysis may result in a poor estimate of the metatechnology so that the corresponding results of productivity will not properly reflect the performance. In order to overcome this problem, the application of the proposed non-convex metafrontier Malmquist index is suggested.

The existing convex and the proposed non-convex metafrontier Malmquist index have been formulated on the basis of the axioms already outlined in Sections 5.2 and 5.4, respectively. These axioms lead the DEA models to be under VRS assumption in both approaches. The reason is that when the inputs (i.e., FTE and WRT) are increased by a certain factor, one cannot necessarily assume that the outputs (i.e., NOC and THT) will also increase by the same factor. The assumption of VRS also ensures that maintenance units are only benchmarked against units of a similar size. This is a property which was also demanded by the representatives of KONE to be satisfied in the analysis.

In order to deal with the undesirable output NOC, a linear transformation approach has been applied to its values. Accordingly, each value of this output is multiplied by (–1) and find a proper translation amount to convert the negative data to non-negative data (for more details about this linear transformation see e.g., Seiford and Zhu 2002). Furthermore, the analysis follows an input-oriented perspective as the maintenance groups are expected to minimize their inputs (i.e., FTE and WRT), controlling for their output levels (an overview of standard DEA models and their features can be found in, e.g., Thanassoulis 1997).

5.5.2 Results and discussions

The mathematical programming problems of the convex and non-convex metafrontier Malmquist indices as well as their components have been encoded in AIMMS, version 3.13. Applied to the aforementioned data set, Table 5.1 summarizes the results of the two frameworks. The first Column represents the units and their corresponding group numbers. EC, BPC and Malmquist index refer to the efficiency change, the best practice change and the Malmquist index stemming from both approaches.

In order to have a general picture of the differences between the results of the two approaches in measuring productivity change, the Spearman's rank correlation test has been carried out. This non-parametric test determines the degree to which two numerical variables (e.g., Malmquist index) are monotonically related or associated (for more details about the Spearman test, see e.g., Miah 2016). With respect to the results in Table 5.1 the Spearman's rank correlation coefficient concerning the Malmquist index is 0.905, meaning that there is a high congruity between the rankings in the two approaches. This is an interesting result as the proposed method here does not change entirely the concept and the structure of how the productivity change is measured. Nonetheless, the non-convex approach provides a more accurate set of results compared to the convex approach as will be investigated in the following.

Table 5.1: Results obtained by the existing and the proposed Malmquist index

	Convex metafrontier Malmquist index			*Non-convex metafrontier Malmquist index*		
	EC	BPC	MI	EC	BPC	MI
Unit 1 (G1)	1.004	0.965	0.969	1.004	0.958	0.961
Unit 2 (G1)	1.087	0.971	1.056	1.087	0.968	1.052
Unit 3 (G1)	1.000	1.217	1.217	1.000	1.266	1.266
Unit 4 (G1)	1.000	1.178	1.178	1.000	1.076	1.076
Unit 5 (G1)	1.200	1.152	1.382	1.200	1.230	1.476
Unit 6 (G1)	1.049	1.079	1.131	1.049	1.100	1.154
Unit 7 (G1)	1.000	1.027	1.027	1.000	1.027	1.027
Unit 8 (G1)	1.000	1.000	1.000	1.000	1.000	1.000
Unit 9 (G1)	1.040	1.083	1.127	1.040	1.122	1.167
Unit 10 (G1)	1.000	0.996	0.996	1.000	0.992	0.992
Unit 11 (G1)	0.954	1.084	1.034	0.954	1.089	1.038
Unit 12 (G2)	1.000	0.949	0.949	1.000	0.939	0.939
Unit 13 (G2)	1.063	0.929	0.988	1.063	0.900	0.957
Unit 14 (G2)	1.008	1.164	1.173	1.008	1.096	1.105
Unit 15 (G2)	1.000	0.971	0.971	1.000	0.979	0.979
Unit 16 (G2)	1.063	1.082	1.150	1.063	1.082	1.150
Unit 17 (G2)	1.189	0.966	1.149	1.189	0.966	1.149
Unit 18 (G2)	1.058	1.022	1.082	1.058	1.022	1.082
Unit 19 (G2)	1.000	0.989	0.989	1.000	1.059	1.059
Unit 20 (G2)	0.987	0.966	0.953	0.987	0.966	0.953
Unit 21 (G2)	1.000	1.130	1.130	1.000	1.002	1.002
Unit 22 (G2)	1.190	1.074	1.279	1.190	1.074	1.279
Unit 23 (G2)	1.098	0.879	0.965	1.098	0.879	0.965
Unit 24 (G3)	1.000	1.136	1.136	1.000	1.035	1.035
Unit 25 (G3)	1.000	1.130	1.130	1.000	1.134	1.134
Unit 26 (G3)	1.000	0.958	0.958	1.000	0.998	0.998
Unit 27 (G3)	1.092	0.943	1.030	1.092	0.941	1.028
Unit 28 (G3)	1.249	0.935	1.168	1.249	0.956	1.194
Unit 29 (G3)	1.174	0.798	0.937	1.174	0.790	0.928
Unit 30 (G3)	1.237	0.784	0.970	1.237	0.784	0.970
Unit 31 (G3)	1.149	0.907	1.042	1.149	0.915	1.051
Unit 32 (G4)	1.000	1.115	1.115	1.000	1.000	1.000
Unit 33 (G4)	1.061	1.055	1.120	1.061	1.000	1.061
Unit 34 (G4)	1.374	0.733	1.007	1.374	0.733	1.007
Unit 35 (G4)	1.000	0.970	0.970	1.000	0.970	0.970
Unit 36 (G4)	0.986	0.848	0.836	0.986	0.890	0.877
Unit 37 (G4)	1.000	1.000	1.000	1.000	1.000	1.000
Unit 38 (G4)	1.053	1.039	1.094	1.053	1.028	1.082
Unit 39 (G4)	1.212	0.792	0.960	1.212	0.774	0.938
Unit 40 (G4)	1.079	0.950	1.026	1.079	1.000	1.079
Unit 41 (G4)	0.964	0.973	0.939	0.964	1.000	0.964

As can be seen in Table 5.1, the results of productivity change obtained by the two approaches diverge substantially for the majority of units. From 41 maintenance units in the data set, only twelve units (7, 8, 16, 17, 18, 20, 22, 23, 30, 34, 35 and 37) yield the same numerical values of the Malmquist index. For the other units, significant differences can be observed. Take unit #19 as an example: Its numerical value of the Malmquist index differs by around 7 % – while the proposed approach captures a positive change in productivity over time (i.e., 5.9 %), the convex metafrontier suggests a decline of 1.1 %.

From a theoretical point of view, this example gives interesting evidence of how a more accurate estimation of the metatechnology leads to a significant difference in the results. Taking a closer look at them, one can also verify which estimated values of productivity change represent properly the performance of unit #19 from a practical point of view: The manager of this maintenance unit has undertaken various efforts (e.g., optimized route planning) to cope with an additional workload (i.e., a higher THT value) while improving its work quality (i.e., reducing its NOC value). Therefore, a productivity decline of 1.1 % (as it is determined by the convex Malmquist index) is indeed counter-intuitive. In contrast, the productivity improvement of 5.9 % attested by the non-convex Malmquist index corresponds closely to the practical expectations of KONE’s management.

As another example, the productivity change of unit #32 amounts to +11.5 % with the convex Malmquist index, while there is no change shown by the non-convex Malmquist index. Analyzing the detailed results, one can observe that the convex approach includes different group technologies from both periods of time to measure the Malmquist index of this unit (i.e., unit #36 from period 1 and units #21 and #32 from period 2). This result (regardless of the value captured) is not readily acceptable by the management due to differences in group technologies over time (see Section 5.5.1 for a few examples of these changes). Furthermore, this combination has led to a very large value of the Malmquist index of +11.5 %, which has also been recognized as a value far away beyond management’s expectations of the performance of this unit. Unit #32 has the reputation of being one of the best performing maintenance units in the whole sample so that its performance is expected to be very high in both observed periods. However, only the non-convex approach has captured a full meta-efficiency of this unit in both periods, resulting in a Malmquist index of one. In contrast, the convex approach estimated a full meta-efficiency in the second period, but a meta-inefficiency in the first period, leading to a large unexpected positive change in productivity over time. This substantial productivity improvement of +11.5 % has been considered as unrealistic by KONE’s management.

The results in Table 5.1 also show that the EC component of the Malmquist index is identical for all units determined by the two approaches. The reason is that both the convex and non-convex metafrontier Malmquist index apply the same ratio of efficiencies to capture the EC component (see formula (5.14) in Section 5.4). This leads to the conclusion that the discrepancies between the results of the Malmquist index originate exclusively from the different estimations of the metatechnology required for the computation of the respective BPC component. As Table 5.1 reports, the BPCs of some units in the existing framework are less than those in the proposed approach, while the opposite is true for some other units. As a key factor affecting the results of productivity change, the BPC component indicates a productivity loss or gain for the majority of maintenance units. Take again unit #19 as an example. As can be seen, a poor estimation of the technology and the corresponding result of the BPC suggest that the productivity of this unit has declined within the existing approach of the Malmquist index, while the enhanced method of estimation within the proposed approach identifies an opposite direction. This underlines the serious drawback of the convex metafrontier Malmquist index concerning the estimation of the benchmark technology and the resulting productivity values, which may lead obviously to wrong conclusions and policy recommendations.

A further diagnosis of this drawback not only can provide specific reasons behind any change in the Malmquist index in the two approaches but also highlight other advantages of the proposed approach. For the determination of the Malmquist index of a maintenance unit *p*, it is required to compute the unit's meta-efficiency against the frontier of the convex and non-convex metatechnologies in both time periods (2014 and 2015), i.e., $Eff^{M}(U_p^{2015}) / Eff^{M}(U_p^{2014})$ in the convex form (see formula (5.4) in Section 5.2) and $Eff_{NC}^{M}(U_p^{2015}) / Eff_{NC}^{M}(U_p^{2014})$ in the proposed non-convex form (see formula (5.13) in Section 5.4). These meta-efficiency scores $Eff^{M}(U_p^{2014})$, $Eff^{M}(U_p^{2015})$, $Eff_{NC}^{M}(U_p^{2014})$ and $Eff_{NC}^{M}(U_p^{2015})$ are reported in the second, fourth, sixth and eighth Columns of Table 5.2, respectively. Moreover, the corresponding reference technologies involved in the computation of these efficiencies are also given next to the efficiency scores in this table.

Table 5.2: Meta-efficiencies and reference technologies in the two approaches

	Convex metafrontier Malmquist index				*Non-convex metafrontier Malmquist index*			
	Period 1		Period 2		Period 1		Period 2	
	Eff	Ref. Technology	Eff	Ref. Technology	Eff	Ref. Technology	Eff	Ref. Technology
Unit 1 (G1)	0.585	P1-G4, P2-G4	0.567	P1-G4	0.590	P1-G4	0.567	P1-G4
Unit 2 (G1)	0.555	P1-G4, P2-G4	0.586	P1-G4, P2-G4	0.559	P1-G4	0.589	P1-G4
Unit 3 (G1)	0.771	P1-G4, P2-G4	0.939	P1-G4, P2-G4	0.790	P1-G4	1.000	P2-G1
Unit 4 (G1)	0.849	P1-G4, P2-G2	1.000	P2-G1	0.929	P2-G2	1.000	P2-G1
Unit 5 (G1)	0.635	P1-G4	0.878	P1-G4, P2-G2	0.635	P1-G4	0.937	P2-G2
Unit 6 (G1)	0.616	P1-G4	0.696	P1-G4, P2-G2	0.616	P1-G4	0.711	P1-G4
Unit 7 (G1)	0.823	P1-G4	0.845	P1-G4	0.823	P1-G4	0.845	P1-G4
Unit 8 (G1)	0.950	P1-G4	0.950	P1-G4	0.950	P1-G4	0.950	P1-G4
Unit 9 (G1)	0.674	P1-G4	0.759	P1-G4, P2-G2	0.674	P1-G4	0.787	P1-G4
Unit 10 (G1)	0.635	P1-G4, P2-G4	0.632	P1-G4, P2-G4	0.648	P1-G4	0.642	P1-G4
Unit 11 (G1)	0.650	P1-G4, P2-G4	0.672	P1-G4, P2-G4	0.658	P1-G4	0.683	P1-G4
Unit 12 (G2)	0.741	P1-G4, P2-G4	0.703	P1-G4, P2-G4	0.792	P1-G4	0.744	P1-G4
Unit 13 (G2)	0.684	P1-G4, P2-G4	0.676	P1-G4, P2-G4	0.718	P1-G4	0.688	P1-G4
Unit 14 (G2)	0.775	P1-G4, P2-G4	0.909	P1-G4, P2-G4	0.884	P2-G4	0.976	P2-G4
Unit 15 (G2)	0.827	P1-G4, P2-G2, P2-G4	0.803	P1-G4, P2-G2	0.923	P2-G4	0.904	P2-G4
Unit 16 (G2)	0.529	P1-G4	0.609	P1-G4	0.529	P1-G4	0.609	P1-G4
Unit 17 (G2)	0.481	P1-G4	0.552	P1-G4	0.481	P1-G4	0.552	P1-G4
Unit 18 (G2)	0.563	P1-G4	0.609	P1-G4	0.563	P1-G4	0.609	P1-G4
Unit 19 (G2)	0.705	P1-G4, P2-G4	0.697	P1-G4, P2-G4	0.719	P2-G4	0.761	P1-G4
Unit 20 (G2)	0.768	P1-G4	0.732	P1-G4, P2-G4	0.768	P1-G4	0.732	P1-G4
Unit 21 (G2)	0.885	P1-G4, P2-G2	1.000	P2-G2	0.998	P1-G4	1.000	P2-G2
Unit 22 (G2)	0.640	P1-G4	0.818	P1-G4	0.640	P1-G4	0.818	P1-G4
Unit 23 (G2)	0.452	P1-G4	0.437	P1-G4	0.452	P1-G4	0.437	P1-G4
Unit 24 (G3)	0.673	P1-G4, P2-G2	0.764	P1-G4, P2-G2	0.748	P2-G1	0.774	P2-G2
Unit 25 (G3)	0.764	P1-G4	0.864	P1-G4, P2-G4	0.764	P1-G4	0.867	P1-G4
Unit 26 (G3)	0.730	P1-G4, P2-G4	0.699	P1-G4, P2-G4	0.749	P1-G4	0.747	P1-G4
Unit 27 (G3)	0.475	P1-G4, P2-G4	0.490	P1-G4	0.476	P1-G4	0.490	P1-G4
Unit 28 (G3)	0.515	P1-G4, P2-G4	0.601	P1-G4, P2-G4	0.515	P1-G4	0.615	P1-G4
Unit 29 (G3)	0.501	P1-G4, P2-G4	0.470	P1-G4	0.506	P1-G4	0.470	P1-G4
Unit 30 (G3)	0.536	P1-G4	0.520	P1-G4	0.536	P1-G4	0.520	P1-G4
Unit 31 (G3)	0.558	P1-G4, P2-G4	0.581	P1-G4, P2-G4	0.564	P1-G4	0.593	P1-G4
Unit 32 (G4)	0.897	P1-G4, P2-G2, P2-G4	1.000	P2-G4	1.000	P1-G4	1.000	P2-G4
Unit 33 (G4)	0.893	P1-G4, P2-G2	1.000	P2-G4	0.942	P1-G4	1.000	P2-G4
Unit 34 (G4)	0.541	P1-G4	0.545	P1-G4	0.541	P1-G4	0.545	P1-G4
Unit 35 (G4)	1.000	P1-G4	0.970	P1-G4, P2-G4	1.000	P1-G4	0.970	P1-G4
Unit 36 (G4)	1.000	P1-G4	0.836	P1-G4, P2-G4	1.000	P1-G4	0.877	P1-G4
Unit 37 (G4)	1.000	P1-G4	1.000	P2-G4	1.000	P1-G4	1.000	P2-G4
Unit 38 (G4)	0.484	P1-G4, P2-G4	0.529	P1-G4, P2-G4	0.512	P2-G4	0.554	P1-G4
Unit 39 (G4)	0.625	P1-G4, P2-G4	0.600	P1-G4	0.639	P1-G4	0.600	P1-G4
Unit 40 (G4)	0.829	P1-G4, P2-G4	0.851	P1-G4, P2-G4	0.927	P1-G4	1.000	P2-G4
Unit 41 (G4)	0.713	P1-G4, P2-G4	0.669	P1-G4, P2-G4	0.717	P1-G4	0.691	P2-G4

As can be seen in Table 5.2, the efficiency values determined by the convex metafrontier Malmquist index are less than or equal the respective efficiencies computed by the proposed approach. This derives from the fact that the non-convex PPS is a subset of the convex technology set (see Section 5.4). When different reference group technologies are applied, discrepancies between the efficiency scores arise. The same efficiency values, however, can be observed where the group technologies are identical under the convex and non-convex metafrontier approaches.

Let us take unit #22 as an example, which has received the same efficiency scores in both approaches (0.640 and 0.818 in 2014 and 2015, respectively). The reason is that the two frameworks have used the same reference group technology P1-G4 (i.e., group technology #4 in period 2014) as well as the same benchmarking peers for this unit (i.e., units #35 and #37). This can mainly be traced back to the size of this maintenance unit. Unit #22 is one of the smallest units in the whole data set in terms of both the THT and FTE values. Since the VRS-specification seeks for benchmarks of a similar size for this unit, only a few comparable maintenance units remain. This increases the probability that both approaches identify the same references, leading also to the same meta-efficiency scores. Investigating the results for this unit, one can see that these identified references are also comparable regarding their other characteristics. For example, unit #22 and also its peers (units #35 and #37) have in common that they are located in big German cities, experiencing a similar business environment. Discussing this result with KONE, it has been confirmed that this precise way of selecting peers is recognized as a powerful feature of the proposed approach by management.

As the results in Table 5.2 show, the convex metafrontier Malmquist index often uses a combination of distinct group technologies from different time periods for the determination of the meta-efficiencies. For example, $Eff^{M}(U_{15}^{2014})$ is based on the reference technologies P1-G4, P2-G2 and P2-G4, i.e., the local technology of group #4 in the first and second period together with group #2 in the second period. However, evaluating units on the basis of such combinations of local technologies seems counter-intuitive from a practical point of view. The application of the convex metafrontier Malmquist index explicitly accepts that all observations – regardless of their respective groups or time periods – can form the metatechnology. Therefore, not only for this unit, but also for all other units in different time periods, this framework does not distinguish between observations which are originated from different local technologies. In other words, observations influenced by a different internal and external environment have constructed together a metatechnology to measure the meta-efficiencies. According to the primary setting of the problem – which suggests grouping of units – and also the fact that the technology has changed over time (see again Section 5.5.1 for a few examples of reasons behind changes in the technology), one can conclude that such combinations of the local technologies cannot be accepted as an accurate way of measuring efficiency.

By contrast, the proposed Malmquist index is immune to this problem. As can be observed in Table 5.2, the proposed approach uses, e.g., solely P2-G4 as a reference technology for measuring $Eff_{NC}^{M}(U_{15}^{2014})$. In other words, the non-convex approach does not make use of combinations of different technologies in the determination of the meta-efficiencies. This property not only leads to a more accurate estimate of the metatechnology but also preserves the characteristics of the local technologies in the form of the best experienced technology over time. This unique feature of the proposed approach plays a crucial role for managing the groups of maintenance units, where, e.g., improving their management styles and promoting corporate learning between groups are sought. For example, the proposed approach identifies P2-G4 as an appropriate benchmark reference technology for unit #15 and some other units. This information can serve KONE as a starting point for a more detailed analysis of P2-G4. Intra-organizational learning seminars, e.g., can be used to analyze the special characteristics of this local group technology, which can be subsequently tested for their applicability to other maintenance units or groups in general and for unit #15 in particular. This example shows how the new approach can substantially support the corporate learning inside KONE and, hence, serve as an additional measure to gradually improve the productivity of the company.

5.6 Conclusions and outlook on future research opportunities

In this chapter, a new way of estimating the metatechnology has been introduced. This approach only applies the minimum extrapolation principle on the aggregation of the experienced group technologies over time. It has been shown theoretically as well as numerically that the resulting new metafrontier Malmquist index, called the non-convex metafrontier Malmquist index, provides more accurate results compared to the existing metafrontier Malmquist index. The proposed index also preserves the role of each group technology – observed at a specific time period – in the estimation of the metatechnology.

This includes information about the superior group technologies observed over time. As exemplified for the case of KONE, this unique feature of the suggested approach can play a crucial role in measuring and analyzing productivity, where a further diagnosis of individual performances is required. With respect to both computational and test properties, the proposed index also possesses the circularity property and is immune to infeasibility under VRS. Similar to traditional indices, it has been decomposed into the standard components such as EC and BPC.

From a theoretical point of view, an interesting perspective for future research would be to extend the proposed non-convex metafrontier approach to other DEA-based frameworks which implicitly accept convex combinations across distinct technologies either to measure the performance in a static setting or in a dynamic environment. For example, frequently applied approaches as the DEA window-analysis (e.g., Charnes et al. 1984) or the sequential Malmquist index (e.g., Shestalova 2003) use a structure which is quite similar to the conventional metafrontier Malmquist index. From a practical point of view, future research should concentrate on analyzing the dual role of maintenance units as both providers of services to the customers and operating units which should contribute to the profit of the KONE Corporation as a whole. Therefore, the performance of KONE's maintenance units regarding financial objectives should be compared with its performance regarding operating objectives.

6 A non-convex metafrontier Malmquist index for group performance comparison[71]

6.1 Introduction

In many settings of practical interest where DEA models are applied, DMUs are partitioned into a few groups, each of which is with the same technology resulting from, e.g., the same resource, regulatory or other managerial and environmental constraints (O'Donnell et al. 2008, Afsharian and Podinovski 2018). Examples for such groupings are categories of schools applying different educational concepts (see e.g., Charnes et al. 1981), dairy farms operating in different countries (see e.g., Latruffe et al. 2012) or local branches of a company organized into groups according to their management styles (see e.g., Huang et al. 2013). In such situations, not only the efficiency of DMUs within their groups but also the comparative performance of the groups of units relative to each other has always been of great interest (see also, e.g., Cook et al. 1998, Cook and Zhu 2007, Cook et al. 2017).

For applications in which comparing the performance of groups of operating units is the fundamental goal, Camanho and Dyson (2006) have developed an index whose structure is built upon the Malmquist index of Färe et al. (1992a). This performance index, however, does not measure the productivity change over a number of time periods but provides a cross-sectional comparison of the performance of groups of DMUs in a static

[71] A slightly modified version of this chapter has been published as Afsharian, M., H. Ahn, S. G. Harms. 2019b. Performance comparison of management groups under centralised management. *European Journal of Operational Research.* Vol. 278(3), pp. 845-854.

setting. An interesting feature of this approach is that the result of the corresponding performance index can be decomposed into various components such as the EI and FI. If the input prices are available, this index can also be decomposed further to capture the root sources of differences in the performance of the groups in terms of cost productivity (see Thanassoulis et al. 2015 and Walheer 2018b). It has also been extended for situations where a pseudo-panel database is available (see Aparicio et al. 2017) and where a baseline group is chosen as reference technology (see Aparicio and Santín 2018).

Among different scenarios where the depicted performance index may be employed, this chapter addresses the case that a central body manages a large number of similar units through a few distinct management groups. Each group – with a segregated geographical business area – has its own unique style in managing its operating units while allocating resources and providing products/services to local customers. It is further assumed that the central management aims to minimize the overall input consumption by the units in each group given the aggregated outputs they produce.[72] Examples of such central management scenarios concern organizations with operating units like bank branches, pharmacy stores, university departments, and police stations, which are often partitioned into a few distinct groups according to their geographical business areas. If the approach of Camanho and Dyson (2006) were to be used under these circumstances, the result might not be optimal, as outlined in the following.

The application procedure of the method of Camanho and Dyson (2006) comprises two key steps within the framework of the Malmquist index. In the first step, group-specific technologies are used and appropriate DEA models – such as the CCR model of Charnes et al. (1978) – are applied to measure the individual efficiency of the operating units. In the following, the performance of each group is captured by an average of the corresponding individual efficiency scores of its operating units, i.e., a post-processing analysis is conducted. As will be shown later in greater detail, by using such an aggregation of the results, this approach implicitly assumes that all operating units in the organization perform independently and pursue their own interest. An individual unit – regardless of the group it belongs to – is allowed to maximize its own efficiency, which may not be optimal

[72] This assumption about input orientation is not essential: It is straightforward to extend the here presented results to output efficiency.

for its group as a whole in terms of, e.g., resource use relative to outcomes. For example, one may argue that a redistribution of resources among operating units in each group may lead to better performance for the group. This is an important characteristic of the system of comparison, which should not be neglected in the modeling process.

As a starting point, the approach of Camanho and Dyson (2006) is extended by introducing an index for comparing the performance of the groups of operating units under the above-outlined scenario of central management. As will be demonstrated later, the resulting approach does not require applying any average of the individual efficiency scores ex post. It will be capable of capturing directly the performance of the groups on the basis of their internal abilities in transforming inputs to outputs. As in the case of Camanho and Dyson (2006), the new approach also builds on the standard Malmquist index of Färe et al. (1992a). Hence, it provides in its basic form a performance index, which fails the circularity property, i.e., the value of the performance index between two groups G1 and G3 cannot be derived from the values of the performance index between G1 and G2 as well as G2 and G3. Like in the approach of Camanho and Dyson (2006), infeasibilities may also occur if the index is decomposed further under the VRS assumption, e.g., for capturing the effect of scale on performance by following the structure as RD decomposition of the Malmquist index in Ray and Desli (1997).

To address these issues, a novel framework of comparing the performance of groups of units is designed, and new centralized DEA models are formulated. In this approach, the performance of the groups is captured against the frontier of the best experienced overall technology (i.e., the best production possibilities) available in principal to all groups of units in the system. Benefitting from this setting, the proposed index can also highlight the technological gap in regard to this potential metatechnology. To this end, the concept of the metafrontier – proposed by Hayami (1969) and further operationalized by Battese and Rao (2002) – is extended in such a way that the resulting common basis of comparison is a pure union of what is really observed rather than also having additional convex combinations of observations originating from different group technologies. As it will be shown later in greater detail, this (non-convex) way of forming the metafrontier also corresponds with the claim of Camanho and Dyson (2006, p. 36) that a performance index may "… not assume convex combinations of group specific frontiers to be feasible. Specifically, even if group-specific production sets satisfy convexity, there is no reason why the union of these sets should be convex." Most recently, this issue has also be emphasized by Aparicio and Santín (2018).[73]

The rest of this chapter unfolds as follows. In Section 6.2 basic notations and assumptions are given. Section 6.3.1 introduces the suggested standard centralized performance index and its components. In Section 6.3.2, the approach is enhanced so that the performance index is capable of passing the circularity property and avoids infeasibility. Section 6.4 highlights the advantages of the proposed indices and their decompositions by means of an empirical illustration to a data set of KONE Corporation. The chapter concludes with a summary and an outlook on future research opportunities in Section 6.5.

6.2 Notations and assumptions

Suppose there exists a set of *n* DMUs that are partitioned into *G* (*G* > 1) distinct groups. Let each group *g* (*g* = 1,…,*G*) include DMUs $(X_j^g, Y_j^g) \in \Re_+^m \times \Re_+^s,\ j = 1,\ldots,\delta_g$, where $X_j^g = (x_{1j}^g, x_{2j}^g, \ldots, x_{mj}^g)$ and $Y_j^g = (y_{1j}^g, y_{2j}^g, \ldots, y_{sj}^g)$ are non-negative and non-zero vectors of inputs and outputs, respectively. Furthermore, assume that DMUs belonging to the same group *g* operate under the same technology, resulting from, e.g., the same resource, regulatory or other environmental constraints. Hence, each local technology of group *g* can be represented by a PPS (technology set) of feasible input-output combinations. It is assumed here that these local technologies are CRS technologies (Charnes et al. 1978) with the conventional axioms of non-emptiness, free disposability, convexity and ray unboundedness.[74] Hence, one can define each group technology T^g, group *g* (*g* = 1,…,*G*),

73 In order to overcome the issue of circularity, Aparicio and Santín (2018) suggest that a pre-specified baseline group can be used as the reference technology. In contrast to this choice of a common technology, the proposed approach applies a metatechnology which is derived from the data of the best production possibilities, keeping also the initial spirit of the methodology designed by Camanho and Dyson (2006) concerning the non-convexification of observations from different groups.

74 It has been shown that the Malmquist index should be measured by the ratio of CRS distance functions even when the technology exhibits VRS (see e.g., Grifell-Tatjé and Lovell 1995; Ray and Desli 1997; Balk 2001; Lovell 2003). Hence, in the design of the here proposed performance-based Malmquist indices, CRS technologies are consistently applied.

as the set of pairs of vectors $(X^g, Y^g) \in \Re_+^m \times \Re_+^s$ for which there exists a vector $\lambda^g \in \Re_+^{\delta_g}$ such that the following conditions are true:

$$\begin{aligned} x_i^g &\geq \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g, \quad \forall i \\ y_r^g &\leq \sum_{j=1}^{\delta_g} \lambda_j^g y_{rj}^g, \quad \forall r. \end{aligned} \tag{6.1}$$

Let the technical efficiency of a DMU under evaluation with the input-output vector (X_p^k, Y_p^k) in regard to technology T^g be represented by $Eff^g(X_p^k, Y_p^k)$. It follows that the efficiency can be measured by solving the DEA model below:[75]

$$\begin{aligned} & Eff^g(X_p^k, Y_p^k) = min\ \theta \\ s.t.\quad & \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g \leq \theta x_{ip}^k, \quad \forall i \\ & \sum_{j=1}^{\delta_g} \lambda_j^g y_{rj}^g \geq y_{rp}^k, \quad \forall r \\ & \lambda_j^g \geq 0,\ \ \theta\ \textit{free in sign}. \end{aligned} \tag{6.2}$$

In model (6.2), λ_j^g stands for intensity variables, θ represents the relative efficiency scores of the unit under evaluation, and the third constraint incorporates the CRS assumption. Furthermore, (6.2) reflects an input-oriented DEA model in the sense that targets are sought that minimize inputs, controlling for output levels. The model can be readily converted to an output-oriented one (see e.g., Thanassoulis 2001).

75 Note that Camanho and Dyson (2006) use distance functions as a representation of technical efficiency. For the sake of simplification, efficiency terms – $Eff^g(X_p^k, Y_p^k)$ – are constantly used throughout this chapter.

6.3 The proposed approach

6.3.1 The standard centralized performance index

Let us revisit the approach of Camanho and Dyson (2006) outlined in Section 3.4.3 in conjunction with the DEA model in (6.2). In each group g (g = 1,…,G), model (6.2) has to be run δ_g times – one after another – in order to determine the efficiency scores of the units. This procedure implies that each unit – regardless of the group it belongs to – may follow an individual way to economize its inputs given the level of outputs it produces for maximizing its efficiency. This flexibility reflects an important property of standard DEA models when there is no grouping of units and DMUs operate independently, each one according to its own priorities. However, this is not a desired approach in the centralized management systems addressed in this thesis.

Recall, that in the here considered scenario a central management aims to minimize the overall input consumption by the units in each group given the aggregated outputs they produce. Hence, the objective is not only improving the performance of the entire system but also preserving the consistency across the units in each group, which is an essential component of the system of comparison. This reveals that the application of the index of Camanho and Dyson (2006) (as showed in (3.26) and its decomposition in (3.27)) is inadequate to compare the performance of the groups under such circumstances. For example, one may argue that a redistribution of resources among operating units in each group may lead to better performance for the group. This is an important characteristic of the system of comparison that should not be neglected in the modeling process.

Against this background and as a “preliminary solution”, it is suggested that the following model – whose original form was proposed first by Lozano and Villa (2004) and further mathematically simplified by Mar-Molinero et al. (2014) – be applied to capture the performance of each group q (q = 1,…,G) against the frontier of group g (g = 1,…,G):

$$\begin{aligned} & Eff_{CM}^{g}(Gr^{q}) = min\ \theta \\ & s.t. \qquad \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g \le \theta \sum_{j=1}^{\delta_q} x_{ij}^q, \qquad \forall i \\ & \qquad\qquad \sum_{j=1}^{\delta_g} \lambda_j^g y_{rj}^g \ge \sum_{j=1}^{\delta_q} y_{rj}^q, \qquad \forall r \\ & \qquad\qquad \lambda_j^g \ge 0,\ \ \theta\ \ free\ in\ sign \end{aligned} \tag{6.3}$$

where the subscript *CM* indicates that the within-group efficiency $Eff_{CM}^{g}(Gr^{q})$ is computed under centralized management. This model is a modification of the DEA model in (6.2) in the sense that for each group *q*, the program now seeks to reduce the total amount of inputs while producing at least the current amount of outputs. Therefore, the central management can examine to what extent the current allocation of resources in group *q* is efficient against the group technology *g* as a benchmark. As this model captures the efficiency of all units in each group *q* in a joint manner, it provides a single efficiency score for the entire group. This is a crucial difference from the conventional DEA model in (6.2), in which the operating units are considered independent and an individual efficiency score is computed for each unit in the system.

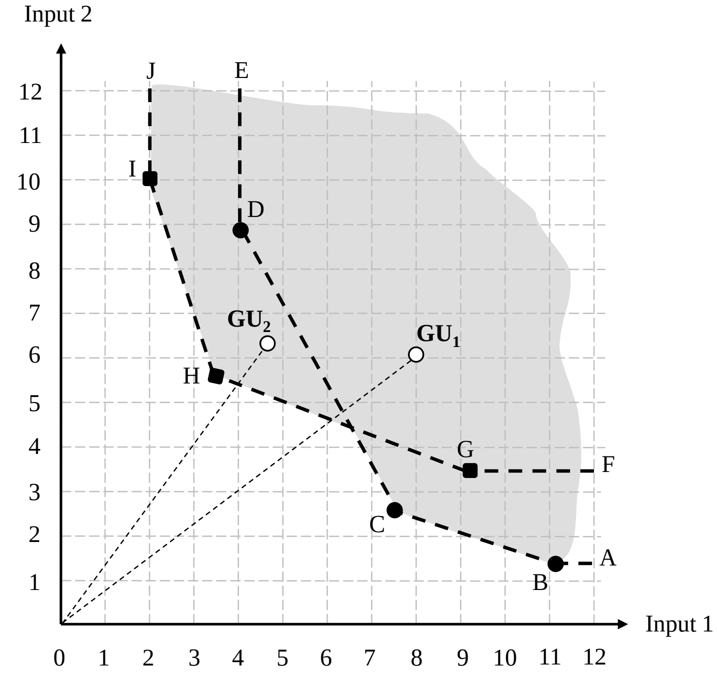
To compare the performance of, e.g., two groups G1 and G2, it is suggested that the index in (3.26) be modified as follows:

$$CPI_{2,1}^{S} = \left[\frac{Eff_{CM}^{2}(Gr^{2})}{Eff_{CM}^{2}(Gr^{1})} \times \frac{Eff_{CM}^{1}(Gr^{2})}{Eff_{CM}^{1}(Gr^{1})} \right]^{1/2} \tag{6.4}$$

Similar to the conventional performance index in (3.26), this CPI^{S} (i.e., standard centralized performance index) is also built on the Malmquist index of Färe et al. (1992a). However, it is equipped by the concept of centralized management, as graphically illustrated in Figure 6.1. In this example, it is supposed that there exist two group technologies T^{1} and T^{2} with two inputs and a single output. As the single output is assumed to be the same for all DMUs, it is not shown. Hence, the group technologies are depicted as the areas bounded by ABCDE and FGHIJ, respectively.

As has been stated by Lozano and Villa (2004) and also formally shown by Asmild et al. (2009), the model in (6.3) can equivalently be interpreted as the measure of performance for a virtual unit that possesses the mean value of inputs and outputs computed across all units in group q against the frontier of group g.[76] This particular interpretation can help to better visualize the way the proposed performance index in (6.4) functions.

Figure 6.1: Two group technologies



According to this example, assume that the corresponding virtual units representing the two groups are depicted as GU_1 and GU_2 in Figure 6.1. In formula (6.4), $Eff_{CM}^{1}(Gr^{1})$ captures the efficiency of the virtual unit GU_1 in respect to its own group frontier (i.e., the border shown by ABCDE). $Eff_{CM}^{2}(Gr^{2})$ can be interpreted similarly for the second group. $Eff_{CM}^{1}(Gr^{2})$ and $Eff_{CM}^{2}(Gr^{1})$ represent the efficiency scores of GU_1 and GU_2 in respect to the frontier of their opposite group, respectively.

Therefore, using group technology T^{2} as a reference, the first ratio in formula (6.4) compares $Eff_{CM}^{2}(Gr^{2})$ with $Eff_{CM}^{2}(Gr^{1})$. The second ratio inside the brackets evaluates the

[76] To see this, we divide both sides of the constraints in (6.3) by δ_q and replace λ_j^g by a new non-negative variable $\widehat{\lambda}_j^g = \lambda_j^g / \delta_q$.

same with reference to group technology T^1. On this basis, a value of $CPI_{2,1}^S$ greater than one implies that the performance of the second group is higher than the first group. The opposite occurs when the index signals the other direction. The two groups will also be evaluated "similar" in performance, if the result of the index is 1.

The proposed performance index can also be decomposed into the following two components:

$$CPI_{2,1}^S = \underbrace{\frac{Eff_{CM}^2(Gr^2)}{Eff_{CM}^1(Gr^1)}}_{Efficiency\ \ Index} \times \underbrace{\left[\frac{Eff_{CM}^1(Gr^2)}{Eff_{CM}^2(Gr^2)} \times \frac{Eff_{CM}^1(Gr^1)}{Eff_{CM}^2(Gr^1)}\right]^{1/2}}_{Frontier\ \ Index}. \tag{6.5}$$

These components can be interpreted analogously to (3.27): the EI compares the relative efficiency spread of the groups, and the FI examines which of these groups has a superior technology.

It is worth comparing the characteristics of the index proposed above with the one of Camanho and Dyson (2006) in greater detail. Note the index in (6.4) and its components in (6.5) apply appropriate adaptations of the centralized DEA model in (6.3). Instances of this model are capable of automatically yielding a single performance score for each group under evaluation. Therefore, in contrast to the conventional index in (3.26), CPI^S does not need to resort to any average to aggregate individual performances, which may be seen as an artificial agglomeration of the DMU-specific performance scores. This is an important feature of the proposed index, which is also kept when the approach will be enhanced in the next section.

The proposed performance index – like the one developed by Camanho and Dyson (2006) – might, however, be infeasible when it is computed and decomposed under the VRS assumption. As an example, following the RD decomposition of the Malmquist index proposed by Ray and Desli (1997), one may extend the decomposition in (6.5) to also comparing the groups concerning their scale efficiency. Since in the RD decomposition, the benchmark technologies are VRS technologies, cross-references $Eff_{CM}^1(Gr^2)$ and $Eff_{CM}^2(Gr^1)$ in (6.5), – like cross-references $Eff^1(X_j^2, Y_j^2)$ and $Eff^2(X_j^1, Y_j^1)$ in (3.26) – might become infeasible. In the approach of Camanho and Dyson (2006), this occurs when even one of the DMUs in a group under evaluation is not enveloped by the frontier formed by the DMUs of the opposite group. This occurrence – although theoretically possible – should, however, be rare in the centralized approach. The reason is that the model in (6.3) determines a single efficiency score for "the group of units" or equivalently for a virtual unit that possesses the mean value of inputs and outputs computed across all units in the group (see e.g., Lozano and Villa 2004, Asmild et al. 2009, Mar-Molinero et al. 2014). Therefore, this virtual unit, which has the average level of each input and each output observed in a group, should normally be enveloped by the frontier of the opposite group.

Another serious drawback with performance indices that use a geometric mean in their structure – including the preliminary approach and the one of Camanho and Dyson (2006) – is that they fail to pass the circularity property. For example, the value of $CPI_{3,1}^{S}$ between the two groups G1 and G3 cannot be indirectly derived from the performance values between G1 and G2 (i.e., $CPI_{2,1}^{S}$) as well as between G2 and G3 ($CPI_{3,2}^{S}$). In other words, $CPI_{3,1}^{S}$ might not be equal to $CPI_{3,2}^{S} \times CPI_{2,1}^{S}$. The circularity property allows the analyst not only to compare the performance of any number of groups in the system but also provides a consistent ranking of their corresponding performances. This is of particular importance in practice, where, e.g., the central management wishes to rank the technologies and determine which group technology is superior to others for planning purposes and further improving the performance of the entire system.

6.3.2 The non-convex metafrontier centralized performance index

To cater for circularity and avoid infeasibilities under a VRS decomposition, it is suggested here that a common reference technology is applied as a base of comparison. As recently discussed by Aparicio and Santín (2018), to overcome the issue of non-circularity, one may apply either a baseline group or a metatechnology. While the former alternative of determining a common reference technology corresponds to a fixed (pre-selected) reference of comparison (i.e., the frontier of one of the groups is chosen by the researcher), the latter results from the envelope of all the units from all groups.[77]

A particular advantage of the fixed reference technology is that the respective performance index (like the one developed by Aparicio and Santín 2018) does not have to assume convex combinations of group specific frontiers to be feasible. In this sense, such an approach shares the same property as in the original method of Camanho and Dyson (2006). This appealing feature is lost, if an ordinary metatechnology – which simply pools the observations – is formed as a reference of comparison. Nevertheless, an advantage of the metatechnology is that it is immune to infeasibilities if the resulting performance index is decomposed further under VRS (e.g., by following the RD decomposition of the Malmquist index, discussed in Section 6.3.1). However, the property of avoiding infeasibilities under VRS decomposition may not be satisfied if the fixed reference technology is taken into account. Since the choice of a fixed reference of comparison is also done subjectively by the analyst, the final results also naturally reflect this choice in the analysis.

Against this background, an approach is proposed in the following which possesses both advantages of the aforementioned two alternatives. In particular, a metatechnology is suggested which is derived from the data of the best production possibilities, keeping also the initial spirit of the methodology in Camanho and Dyson (2006) concerning the non-convexification of observations from different groups. Such a metatechnology can built by adapting and extending the metafrontier methodology proposed by Hayami (1969) as well as Hayami and Ruttan (1970) and further developed by Battese and Rao (2002), Battese et al. (2004) as well as O'Donnell et al. (2008) (see Section 3.3.2). Within this approach, a group technology is formed by considering all units belonging to the same group whereas the metatechnology is the union of the group technologies. Hence, the

[77] The idea of the fixed reference technology (to cater for circularity) can be seen in the structure of the based-period Malmquist index of Berg et al. (1992). Applying a metatechnology as a reference technology (to overcome infeasibilities and satisfy circularity) has also been employed in the global form of the Malmquist index of Pastor and Lovell (2005).

metatechnology for the group of technologies T^g, g (g = 1,…,G) can be defined as follows:[78]

$$T^M = T^1 \cup T^2 \cup,...,\cup T^G. \tag{6.6}$$

This metatechnology represents the best experienced overall technology available in principle to all groups of units in the system. It should be noted here that the resulting metatechnology is a "pure" union of what is really observed rather than also perhaps having additional areas resulting from any convexication between group technologies.[79] Hence, with this particular definition of the common reference of comparison in (6.6), the proposed approach keeps the property given by Camanho and Dyson (2006) by which it is not necessary to assume that convex combinations of group specific frontiers are feasible.

Taking into account the definition in (6.6), the performance of a group q against the metafrontier of technology T^M can be captured by the following mixed integer linear programming problem:

$$\begin{aligned}
&Eff_{CM}^{M}(Gr^q) = min\ \theta \\
&s.t. \quad \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g \le \theta \sum_{j=1}^{\delta_q} x_{ij}^q + Z\kappa_g, \qquad \forall i, \forall g \\
&\qquad\ \ \sum_{j=1}^{\delta_g} \lambda_j^g y_{rj}^g \ge \sum_{j=1}^{\delta_q} y_{rj}^q - Z\kappa_g, \qquad \forall r, \forall g \\
&\qquad\ \ \sum_{g=1}^{G} \kappa_g = G-1, \\
&\qquad\ \ \lambda_j^g \ge 0,\ \ \kappa_g \in \{0,1\},\ \ \theta\ \textit{free in sign.}
\end{aligned} \tag{6.7}$$

---

78 Studies in which a similar form of the metatechnology has also been applied can be found, e.g., in Tiedemann et al. (2011), Medal-Bartual et al. (2012), Huang et al. (2013), Afsharian et al. (2018a), as well as Walheer (2018a).

79 For studies where a convex form of the metatechnology (i.e., $T^M = conv\{T^1 \cup T^2 \cup,\ldots,\cup T^G\}$) has been applied, see e.g., Kontolaimou and Tsekouras (2010), Oh and Lee (2010), Portela et al. (2011), Fallah-Fini et al. (2012), Zhang et al. (2013) as well as Zhang and Wei (2015).

In model (6.7), the inputs and outputs of the units in all group technologies are initially combined in the first two constraints. However, the binary variables κ_g as well as a sufficiently large constant Z ensure that only one of the groups will be selected at any time. Hence, according to the third constraint, in any feasible solution of (6.7), only one of such binary variables, corresponding to, e.g., $g = g^*$, is equal to 0, while the remaining binary variables are equal to 1. This group $g = g^*$ is the one by which the objective function θ is minimised. Therefore, the central management can examine to what extend the current allocation of resources in group q under evaluation is efficient against the best experienced overall technology available to all groups of units in the system represented by T^M.

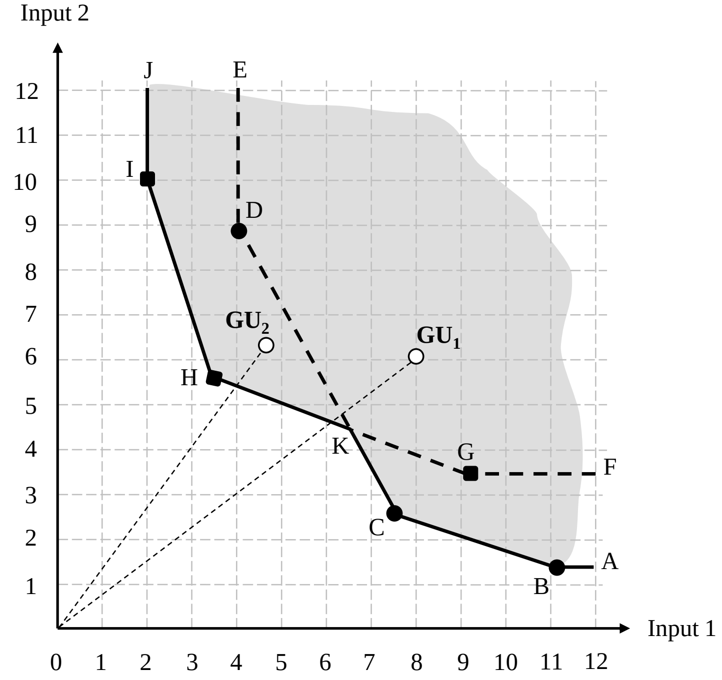
To enhance the definition of the primary standard centralized performance index in (6.5), it is now suggested that a metatechnology is used as a basis of comparison in the following way:

For the comparison of the performance of, e.g., two groups G1 and G2, the group technology T^1 and T^2 in (6.4) are replaced with the metatechnology T^M. Since there is only a single reference of comparison, there is no need to resort to the geometric mean convention when defining the performance index. Hence, the <u>m</u>etafrontier <u>c</u>entralized <u>p</u>erformance <u>i</u>ndex (i.e., CPI^M) is defined by means of the following ratio:

$$CPI_{2,1}^{M} = \frac{Eff_{CM}^{M}(Gr^2)}{Eff_{CM}^{M}(Gr^1)}. \tag{6.8}$$

Turning to the example introduced in Section 6.3.1, Figure 6.2 depicts now the metatechnology of the two group technologies T^1 and T^2. This metatechnology is the area bounded from below by ABCKHIJ.

Figure 6.2: Metatechnology of the two group technologies



According to the just given example, in formula (6.8), $Eff_{CM}^{M}(Gr^{1})$ computes the efficiency of the virtual unit GU_1 in respect to the metafrontier (i.e., the border shown by ABCKHIJ). $Eff_{CM}^{M}(Gr^{2})$ can be interpreted similarly for the second group. Hence, using metatechnology T^{M} as a single reference of comparison, the ratio in formula (6.8) measures the performance of units in group G2 compared to the performance of units in group G1. The greater the ratio is, the higher is the performance of the second group compared to the first group, and vice versa. When this ratio equals 1, then the performance of these two groups are similar.

The performance index in (6.8) can also be represented by the following decomposition:

$$CPI_{2,1}^{M} = \underbrace{\frac{Eff_{CM}^{2}(Gr^{2})}{Eff_{CM}^{1}(Gr^{1})}}_{Efficiency\ Index} \times \underbrace{\left[\frac{Eff_{CM}^{M}(Gr^{2})}{Eff_{CM}^{2}(Gr^{2})} \times \frac{Eff_{CM}^{1}(Gr^{1})}{Eff_{CM}^{M}(Gr^{1})}\right]}_{Technology\ Gap\ Index}. \tag{6.9}$$

Similar to (6.5), the EI in (6.9) compares the relative efficiency spread of the groups. A value of EI greater than one indicates that DMUs in group G2 are closer to their own frontier than are those DMUs in group G1 relative to their own frontier (and vice versa). In the second component, $Eff_{CM}^{1}(Gr^{1})$ is the within-group efficiency of units in group

G1 measured against their own group frontier. $Eff_{CM}^{2}(Gr^{2})$ can be interpreted similarly for the second group. $Eff_{CM}^{M}(Gr^{2})$ and $Eff_{CM}^{M}(Gr^{1})$ capture, however, the performance of the groups where the frontier of the metatechnology is considered as a reference of comparison. Hence, the technological gap index (TGI) in (6.9) compares the distance between the respective frontier of groups G1 and G2 from the metafrontier.[80] In other words, it indicates whether the frontier of group G2 is closer to or farther away from the metafrontier than is the frontier of group G1. For example, a value of TGI larger than unity indicates that DMUs in group G2 have a superior technology than their respective counterparts in group G1 (and vice versa).

It should be emphasized that the metafrontier centralized performance index in (6.8) – in contrast to the standard centralized performance index in Section 6.3.1 and the one of Camanho and Dyson (2006) – satisfies circularity. To see this, assume that there exist three group technologies T^{1}, T^{2} and T^{3}. It can readily be verified that $CPI_{3,2}^{M} \times CPI_{2,1}^{M} = CPI_{3,1}^{M}$. Similarly, one can simply prove that the components of the index in (6.9) are also circular. The circularity property allows the analyst not only to compare the productivity of any number of groups in the system but also provides a consistent ranking of their corresponding performances. This is of particular importance in practice, where, e.g., the central management wishes to rank the technologies and determine which group technology is superior to others for planning purposes and further improving the performance of the entire system.

Within the standard performance index in Section 6.3.1, an infeasibility can occur when either $Eff_{CM}^{1}(Gr^{2})$ or $Eff_{CM}^{2}(Gr^{2})$ cannot be computed under VRS. For example, in the case of $Eff_{CM}^{1}(Gr^{2})$, it may happen that the virtual unit representing group G2 cannot be enveloped by the boundary formed by the units in group G1. This issue can similarly occur in the approach of Camanho and Dyson (2006) where cross-comparisons are done (see the discussions in Section 2.1). However, the metafrontier performance index in (6.10) avoids

[80] In a broader context, the TGI in (6.9) can be seen as an adaption of the “technology gap ratio” which appears in the context of metafrontier analysis (for a comprehensive discussion, see Walheer (2018a).

such infeasibilities. The reason is that all the comparisons are done against the metafrontier, which is the envelop of all group technologies. Hence, by construction, the virtual unit representing any group is always enveloped by the boundary of the metatechnology

6.4 Empirical illustration

6.4.1 Data set and model specification

To illustrate the proposed approach, a set of KONE´s maintenance units is analyzed. Since only a single year is considered within the analysis (i.e., 2015), there was no need to search for representative data sets in *two* time periods (in contrast to the empirical illustration in Chapter 5). Therefore, a higher number of DMUs could be analyzed. The final data set consists of 56 maintenance units (i.e., n = 56 DMUs). These local maintenance units have been partitioned into four distinct managerial groups.[81] For the sake of data anonymization, a randomly selected number from 1 to 4 is given to each of these groups, to which we refer to as G1 (17 units), G2 (12 units), G3 (15 units) and G4 (12 units) in the following.

For the performance comparison of the four groups of KONE, two inputs and two outputs are used. The inputs are FTE and *material costs (MC)* – i.e., the total costs incurred on procured materials (e.g., spare parts or machine oil). The outputs of the study are *the generated revenue (REV)* and THT. REV is the revenue gained by each unit from maintenance and repair tasks. The modifications of the underlying input-output-model was demanded by the representatives of KONE. However, it is also reasonable to incorporate the indicators MC and REV, since they can be predominantly determined by strategic decisions on the management group level. In contrast to that, the NOC and WRT (as used in the empirical illustration of Chapter 5) are predominantly affected by the daily operation of the subordinated technicians (i.e., DMU level).

[81] See Section 2.1 for more information on KONE’s maintenance units and managerial groups.

The analysis follows an input-oriented perspective, as the DMUs are expected to minimize their inputs, controlling for their output levels. Descriptive statistics of the inputs and outputs for these groups are given in

Table 6.1.

Table 6.1: Summary statistics of the used data set

		Average				Standard deviation			
	δ_g	*FTE*	*MC*	*REV*	*THT*	*FTE*	*MC*	*REV*	*THT*
G1	17	8.76	353,434,59	1,082,694.00	1,040.65	2.34	91,568.53	267,839.46	278.22
G2	12	9.98	324,240.50	987,141.67	1,107.38	1.83	53,040.23	151,620.11	278.64
G3	15	8.61	343,056.93	868,557.80	978.73	2.83	115,454.81	299,916.14	205.08
G4	12	8.59	248,369.17	998,173.75	1,039.08	2.76	94,877.14	332,029.89	263.55

As it has been previously discussed, basic DEA models (such as the one in (6.2)) put the DMUs in their best possible light. This is done by maximizing the individual performance of each operating unit in the system. This particular setting reflects an important property of standard DEA models when there is no grouping of units and DMUs operate independently, each one according to its own priorities. Collaborations between operating units in each group in terms of, e.g., within-group coordination, sharing knowledge, redistributions of resources are only a few examples of what the central management of KONE wishes to promote in its system of maintenance units. However, these important characteristic may not be sufficiently covered if such decentralized DEA models are employed in the design of the performance comparison indices. For this case, the application of the proposed centralized performance indices seems to be appropriate.

6.4.2 Results and discussions

The alternative performance indices as well as their components have been computed by solving the corresponding mathematical programming problems, encoded in AIMMS, version 3.13. Table 6.2 summarizes the results of the standard centralized approach as introduced in Section 6.3.1**.** For the sake of comparison, the results obtained by applying the approach of Camanho and Dyson (2006) are also presented.

Table 6.2: Results of the existing and proposed standard centralized approach

	Approach of Camanho and Dyson (2006)			*Standard centralized approach*		
	EI	FI	*PI*	EI	FI	CPI^{s}
G2-G1	0.943	1.046	0.986	0.968	1.025	0.992
G3-G1	0.912	1.030	0.939	0.889	1.050	0.934
G4-G1	0.981	1.164	1.143	0.992	1.147	1.138
G3-G2	0.967	0.988	0.956	0.918	1.001	0.919
G4-G2	1.042	1.138	1.185	1.026	1.170	1.199
G4-G3	1.076	1.174	1.263	1.116	1.215	1.357

Table 6.2 reports the results of the EI, FI and the corresponding performance indices (i.e., the PI of Camanho and Dyson 2006 and the centralized performance index shown by CPI^{s}). In the following, the "Camanho and Dyson" approach is referred to as the CD approach.

Taking G2-G1 as an example, the EI obtained by the proposed approach is equal to 0.968 (compare it to the respective value by CD approach which is 0.943). According to the discussions in Sections 3.4.3 and 6.3, this means that the efficiency spread in G2 is larger than in G1. In other words, the maintenance units in group G1 are closer to their own frontier than are those units in group G2, relative to their own frontier. This implies that the management group G2 faces more improvement potentials than the management group G1 and, hence, should substantially strive for more performance gains in its group. Turning to the second component as a complementary measure, the FI compares the distance between the respective frontiers of groups G1 and G2. For example, the value of 1.025 for G2-G1 in the proposed approach (1.046 by CD approach) indicates that maintenance units in group G2 benefit from a superior technology compared to their respective counterparts in group G1. The overall performance is computed by CPI^{s} in the new approach, which multiplicatively aggregates the results of the EI and FI. In the case of G2-G1, the corresponding value of 0.992 shows that G1 represents a higher overall performance than G2. The PI of the CD approach also represents the same figure, but with a difference in its value (i.e., PI = 0.986).

Considering the result of the CPI^{s} together with the EI and FI, one can obtain a holistic view of performance differences between G2 and G1 and understand the underlying sources of these differences. Although technology T^{2} is superior to technology T^{1}, the efficiency spreads in G2 are substantially larger than in G1, which leads to a lower overall performance in G2 (as indicated by the CPI^{S}). As a managerial implication in the case of KONE, it follows that a superior technology exists in G2 (compared to G1), which may be interpreted by well-established within-group strategies in this group. However, this superior technology has not already been employed by all the maintenance units in G2 uniformly so that a lower overall performance for this group can be observed compared to G1. Therefore, efficiency gains with respect to the existing technology should be a major focus of future management activities for units in group G2.

Different approaches are often constructed on the basis of different assumptions and follow different objectives. Hence, it is not surprising that the results of alternative approaches turn to be different. On the other hand, any tendency in the results of two even competing approaches may also be linked into the structure of a particular data set at hand. Nevertheless, it is worth mentioning that the results of the performance index and its decomposed components computed by the CD approach – at least on the basis of our data set of KONE – represent almost the same general tendency compared to our approach.[82]

The results of the metafrontier centralized approach are summarized in Table 6.3. The first Column shows the names of the groups in comparison. Columns 2-4 contain the numerical values of the EI, the TGI and the performance index (CPI^{M}), respectively. The last Column also reports the resulting best overall technology, which has been recognized as the reference of comparison in our approach.

Table 6.3: Results of the metafrontier centralized approach

	EI	TGI	CPI^{M}	Reference technology
G2-G1	0.968	0.917	0.880	G4

82 Unlike in the proposed approach, the FI of G3-G2 computed with the CD approach ranks G2 after G3 (i.e., FI = 0.988).

G3-G1	0.889	0.944	0.840	G4
G4-G1	0.992	1.096	1.088	G4
G3-G2	0.918	1.029	0.944	G4
G4-G2	1.026	1.196	1.227	G4
G4-G3	1.116	1.161	1.297	G4

According to the structure of the proposed indices in Sections 6.3.1 and 6.3.2 and the results in Table 6.3, the metafrontier approach as well as the standard centralized approach (see Table 6.2), lead to identical EI values. Taking G2-G1 as an example, it can be numerically verified that both centralized indices yield an EI score of 0.968. This corresponds to a higher within-group efficiency in group G1 compared to group G2. The interpretation and the corresponding managerial implications are similar to what have been given above; hence, they are not repeated again here.

Comparing the results of Tables 6.2 and 6.3, differences between the standard and metafrontier approach can be identified for the FI and TGI, respectively. Nevertheless, most values show similar tendencies, i.e., if the FI is greater than unity in the standard approach, one can observe the same direction for TGI in the metafrontier approach (cf. G4-G1, G3-G2, G4-G2 and G4-G3). An opposite direction, however, exists for the results of G2-G1 and G3-G1. Taking again G2-G1 as an illustrative example, one can see that the standard approach yields a value of FI = 1.025, while the metafrontier approach has captured a value lower than unity, i.e., TGI = 0.917. This means that technology T^2 is recognized to be superior to technology T^1 according to the standard index while the metafrontier approach suggests the opposite. To explain such differences, it should be highlighted here again that the standard approach forms a basis of comparison with only the data from technologies under consideration, i.e., technologies T^1 and T^2. In contrast, the metafrontier approach uses information of all groups to form the benchmark technology, i.e., the pure union of technologies T^1, T^2, T^3 and T^4. Accordingly, the result of the metafrontier framework is seen to be more comprehensive compared to the standard centralized approach in this case.

This result has also been confirmed by the KONE Corporation. From a managerial point of view, discussions with representatives of KONE have showed that the results of the metafrontier approach are much closer to the expectations of the corporate management. In their opinion, e.g., group G1 has the reputation of applying a clear and successful business strategy, which has been implemented by the respective management based on detailed process analysis. This strategy is especially characterized by well-organized working processes that enable the technicians to handle specified maintenance tasks in an effective and balanced time- and cost-oriented way. Therefore, the units of group G1 have been able to meet even stricter budget constraints. The corporate management also believes that this specific structure and the corresponding style of management have also thoroughly been implemented in all the maintenance units of group G1. Hence, a less productive technology T^1 compared to T^2 or T^3 – as captured by the FI of the standard centralized approach – has not been recognized as consistent with representatives' point of view about the performance of these groups.

As it has been discussed in Section 6.3.2, using a common reference technology has the advantage that the performance index and the respective components satisfy circularity. Using the numerical values of G2-G1, G3-G1 and G3-G2 from Table 6.3, one can see that the mathematical relationship $CPI_{2,1}^{M} \times CPI_{3,2}^{M} = CPI_{3,1}^{M}$ holds. A major advantage of satisfying circularity is that it allows for ranking the groups according to their performance values, and this ranking always remains consistent. Using technology T^1 as a "basis of ranking", one receives the following order under the non-convex approach: $CPI_{4,1}^{M} > 1 > CPI_{2,1}^{M} > CPI_{3,1}^{M}$. That means that group G4 reveals the highest performance, followed by groups G1, G2 and G3, respectively. Even though the results of the CPI^{S} suggest the same ranking (as shown in Table 6.3), it should be emphasized again that the standard index does not satisfy circularity. Consequently, this feature of ranking may not be seen in other numerical examples.

As can be observed in Table 6.3, the proposed approach has identified a single group G4 as the common reference technology for evaluating the group performances. It should be noted that in other cases with a different data set, it may happen that a union of more than one group is formed. However, in any case, the union of the technologies remains pure or non-convex in the sense that convex combinations of different technologies are not used in the determination of the overall common reference technology (see Section 6.3.2).

In this way, the characteristics of the local technologies in the form of the best experienced technology are preserved and the corresponding performance index is still in line with the initial spirit of the CD approach concerning the non-convexification between group technologies. Furthermore, this non-convexification of observations from different groups can play a crucial role in identifying superior technologies (such as the technology of group G4) and, finally, promote the overall corporate learning.

Turning to the managerial implications, a possible reason for the better performance of group G4 (compared to the other three groups) has been identified through a subsequent analysis of its operating process. Due to the directives of the business management of group G4, technicians in its region perform their tasks with high quality, which is – according to KONE representatives – one of the critical success factors in delivering maintenance services to customers. In order to do so, the group management has shown to be able to reduce cost-intensive repairs caused by worn machine parts not inspected in the course of regular maintenance routines. Such findings obtained through the application of the presented framework can play a crucial role for managing KONE's groups of maintenance units. For example, KONE can use this information for further process inspections of group G4 and identify other special characteristics of this local group technology, which – in combination with intra-organizational learning processes – help to subsequently applying identified superior maintenance strategies to other maintenance units or groups in general. Hence, the results in this section can serve KONE as a possible starting point to incentivize the maintenance groups to improve performance in future.

At this point, it is also interesting to show the relation between the proposed approach and the method recently introduced by Aparicio and Santín (2018). In the context of a decentralized management, in order to overcome the issue of circularity of the CD approach, they have proposed that a pre-specified baseline group can be used as the reference of comparison (see the discussions in Section 6.3.2). For example, in an empirical illustration, the "author's selection" of the baseline technology rests upon the widely recognized reputation of Finland's secondary education "as one of the best performers in cross-countries evaluations" (see Aparicio and Santín 2018, p. 231). In contrast, the here developed approach compares each group to the best overall technology in the sample, which has turned – in this particular case – to be a single group technology G4. Hence, if one selects (in advance) G4 as the baseline technology, an extended approach of Aparicio and Santín (2018) under centralized management (i.e., by applying an appropriate modification of the proposed centralized DEA models in Section 6.3.2) produces the same results as the approach presented in Table 6.3. However, this cannot be generalized as the here developed methodology may form a metafrontier from the data of the best production possibilities in the sample. This may include a union of more than one group.

6.5 Conclusions and outlook on future research opportunities

This chapter has revisited the structure of the approach of Camanho and Dyson (2006) for performance comparison of management groups under centralized management. Among different scenarios where this index may be employed, the situation has been addressed (motivated by the case of KONE Corporation) where a central body manages a large number of similar units through a few distinct management groups. Each group – with a segregated geographical business area – has its own unique management style in managing its operating units while allocating resources and producing products and/or services to local customers.

It has been discussed that the approach of Camanho and Dyson (2006) may not be appropriate under these circumstances, as their approach assumes that all operating units in the organization perform independently and pursue their own interests. An individual unit – regardless of the group it belongs to – is allowed to maximize its own efficiency, which may not be optimal for its group as a whole in terms of, e.g., resource use relative to outcomes.

Against this background, a preliminary approach has been designed that is capable of capturing directly the performance of the groups on the basis of their internal abilities in transforming inputs to outputs. As in the case of Camanho and Dyson (2006), the proposed performance index, however, fails the circularity property. Furthermore, infeasibilities can still occur when the index is computed and decomposed under the VRS assumption. To address these issues, a novel framework of comparing the productivity of groups of units has been developed, and new centralized DEA models have been formulated. In this approach, the performance of each group is captured against the frontier of the best experienced overall technology available to all groups of units in the system. To form such an overall frontier, the concept of the metafrontier has been extended so that the resulting common basis of comparison is a pure union of what is really observed rather than also having additional convex combinations of observations originating from different group technologies.

From a theoretical point of view and with an empirical application to KONE Corporation, the capabilities of the proposed indices have been illustrated. In particular, it has been shown that the resulting metafrontier centralized performance index is not only circular but can also highlight the technological gap in regard to the potential technology available to each group. An interesting perspective for future research is to extend the proposed indices to dynamic environments where a panel data set is available. In this line, one may benefit from the approaches that have recently been proposed in the literature such as the ones of Aparicio et al. (2017) and Afsharian et al. (2018a). Another direction for future research is to decompose the indices further to capture the root sources of differences in the performance of the groups in terms of, e.g., cost productivity along with the proposals of Walheer (2018a) and Walheer (2018c).

7 Final conclusions

In recent decades, the way a company's performance is measured has changed fundamentally. Until the 1980s, the majority of applied frameworks have evaluated a company's performance in accordance with the main shareholder interest of profit maximization. Due to increased competition on globalized markets, approaches developed after the 1980s also assess a company's ability to meet customers' and employees' requirements (see Kaplan and Norton 1996a). Therefore, the various stakeholder perspectives (e.g., shareholders, customers, employees) are typically reflected through a set of different financial and non-financial performance indicators. There is also a widespread agreement within the scientific literature that performance measurement approaches should be tailored to the respective organizational setting (see e.g., Globerson 1985, Maskell 1991, Neely et al. 1995). This can be justified with the tremendous impact of organizational structures on behavior patterns, decision opportunities and, consequently, improvement potentials.

In order to monitor the performance of so-called maintenance units, KONE Corporation – a widely recognized leader in the elevator and escalator industry – also applies a set of financial and non-financial performance indicators. The different units are responsible for the maintenance and repair of elevators and escalators within their defined geographical regions. In order to oversee its maintenance units in an efficient way, KONE has partitioned them into four distinct managerial groups (e.g., North, East, South and West) whereby each group is organized by a regional manager. The respective manager has the ability to reallocate resources, set binding targets and promote collaborations.

Like any other company, KONE faces different challenges that accompany the application of a traditional performance measurement approach. Among other things, KONE's management must explicitly define a set of weights to aggregate the different indicators to a single overall performance score. Such weightings are not only highly subjective but may also cause numerous discussions with other managers about an appropriate weighting scheme. Furthermore, traditional frameworks typically cannot account for certain organizational influence factors. For example, it is usually problematic to incorporate different returns to scales or additional improvement potentials from centrally coordinated resource allocations within the performance measurement approach. However, both aspects are tremendously important for receiving binding and realistic as well as motivational target values.

Based on the aforementioned examples, it seems questionable that traditional approaches can sufficiently measure performance and identify the full improvement potentials of KONE's maintenance units and groups. In response to the limitations of traditional frameworks, the non-parametric approach of data envelopment analysis (DEA) – originally proposed by Charnes et al. (1978) – has been proven as a useful tool to evaluate the performance of so-called Decision Making Units (DMUs) even when multiple indicators are simultaneously applied. This approach uses a model endogenous weighting to receive a single overall performance score and, therefore, does not require previous determination of weights by management representatives. Besides, DEA can account for numerous contextual factors (such as returns to scales) or the additional improvement potentials from, e.g., centrally coordinated resource reallocations. It is therefore not surprising that the business science literature also considers DEA as "an appropriate method for the quantitative comparison of maintenance organizations" (see Garg and Deshmukh 2006, p. 223). Against this background, it is the fundamental objective of this thesis to develop a DEA-based performance measurement approach to appropriately measure the performance of organizations similar to the case of KONE's local maintenance units and regional groups.

Chapter 2 provides basic information about the focal company KONE Corporation and its German subsidiary KONE GmbH. The chapter also includes a study of the structure and recent developments in the elevator and escalator industry. This research points out that the service segment has been faced with increasing price competition in recent years. Simultaneously, new maintenance concepts and time constraints have caused massive work intensification for technicians. Nevertheless, customers are expecting high service quality. Altogether, these fundamental market changes force KONE and its competitors to continuously improving the performance of their service operations.

Chapter 3 gives a thorough introduction to the theory and mathematical foundations of DEA. The chapter emphasizes how the methodology of DEA can handle major limitations of traditional performance measurement approaches. However, it also shows some noteworthy drawbacks of basic DEA models. These drawbacks include (1) challenges in defining an appropriate reference technology, (2) the low discrimination power of performance scores (especially when the number of variables is high in comparison to the number of observations), (3) extremely high or low weights that are attached to some inputs or outputs and (4) target values that are inconsistent with the respective organizational setting. The chapter also includes an overview of how DMUs operating under different technologies can be compared using the approach proposed by Battese and Rao (2002) and further extended by Battese et al. (2004) as well as O'Donnell et al. (2008). Their approach uses an all-encompassing global benchmark technology – a so-called metatechnology – as a referent for all units under assessment. In most applications, it is assumed that the metatechnology is either a convex or non-convex production frontier. The *convex metatechnology* is based on the fundamental assumption that combinations of local technologies are producible, whereas the *non-convex metatechnology* supposes that an aggregation of distinct technologies is inconsistent with the underlying organizational setting. In the same chapter, the DEA-based Malmquist index as proposed by Färe et al. (1992a) is described as a measure to evaluate a unit's productivity changes over time. For applications in which comparing the performance of entire DMU groups is the fundamental goal, the approach of Camanho and Dyson (2006) is also discussed. This performance index is founded on the idea of Färe et al. (1992a) but provides a cross-sectional comparison of the performance of groups in a static setting.

Chapter 4 starts with a brief introduction of the five most frequently discussed organizational variables: specialization, coordination, centralization, configuration and formalization. In order to receive a manageable amount of publications and due to its special relevance for the case of KONE Corporation, the concept of centralization is discussed in more detail from the perspective of DEA. This discussion includes a description of how different levels of centralization are modeled in DEA. A *complete centralization* forces each DMU to operate according to the preferences of the central decision maker. This is mathematically expressed in DEA through the uniform application of a common set of weights to all entities under consideration. A *complete decentralization* means that each

DMU can operate in a self-reliant way and follow its own preferences. Therefore, corresponding DEA models allow each unit to choose an individual set of weights that most suitably represents the preferences of the respective DMU manager. Being aware that in practical situations one can typically observe neither a complete centralization nor a complete decentralization of decisions, a compromise solution approach called "hybrid DEA" was also defined. In this approach complete weight flexibility and a common set of weights are combined.

Chapter 4 also provides a systematic literature review of how the hybrid and centralized management scenario are already modeled in the current DEA literature. In the end, 135 different approaches were identified. According to the respective objectives of each article, the publications were categorized into eight distinct research streams. The study of these streams showed that they partially correspond to the limitations of basic (i.e., decentralized) DEA approaches. This indicates the appealing features of common or hybrid weighting schemes for avoiding fundamental problems of basic DEA models (such as extreme weights or a low discrimination power). However, two significant research gaps were also identified based on the literature review. The first gap is a direct consequence of the basic assumption of the traditional metafrontier Malmquist index that convex combinations of local technologies are attainable and, hence, producible. This is clearly inappropriate for companies that intentionally separate their operations into different business segments and, therefore, make extensive use of the concept of specialization. For these cases, the traditional metafrontier Malmquist index may yield inaccurate performance scores since distinct groups are implicitly mixed in the mathematical computations. The second research gap is caused by the impossibility of comparing the performance of groups of units in centralized management scenarios. So far, the literature only comprises DEA-based approaches that can compare group performances in decentralized management scenarios (see e.g., Camanho and Dyson 2006).

In response to the first research gap, Chapter 5 proposes an alternative metafrontier Malmquist index for measuring productivity over time, where the panel data comprise groups of units operating under the influence of different local technologies. The suggested approach overcomes the aforementioned weakness of the conventional metafrontier Malmquist index that is caused by the convex aggregation of underlying group technologies. The proposed approach uses a new way of estimating the metatechnology, which only applies the minimum extrapolation principle on the aggregation of the group technologies experienced. As a result, one will usually receive more accurate results compared to the existing metafrontier Malmquist index. The proposed index also preserves the role of each group technology – observed at a specific time period – in the estimation of the metatechnology. This means that particularities of groups (e.g., customized strategies) are not mixed within the modeling process. As exemplified for the case of KONE, this unique feature of the suggested approach can play a crucial role in measuring and analyzing productivity, where a further diagnosis of individual performances is required. With respect to both computational and test properties, the proposed index is immune to infeasibility under variable returns to scale (VRS) and also possesses the circularity property.

In respect to the second research gap, Chapter 6 extends the DEA approach of Camanho and Dyson (2006). A beneficial feature of the index proposed here is that improvement potentials resulting from resource redistributions or group-wide collaborations are explicitly incorporated in the underlying mathematical models. In other words, the new framework is applicable to situations where a central decision maker controls the set of DMUs. Besides, the resulting index and its components not only highlight the technological gap in regard to the potential technology available to each group but also satisfy the circularity property.

In summary, this research contributes three major aspects to the current literature which can be classified according to the performance management circle of Ahn (2003) as follows (see also Figure 7.1 for a graphical overview):

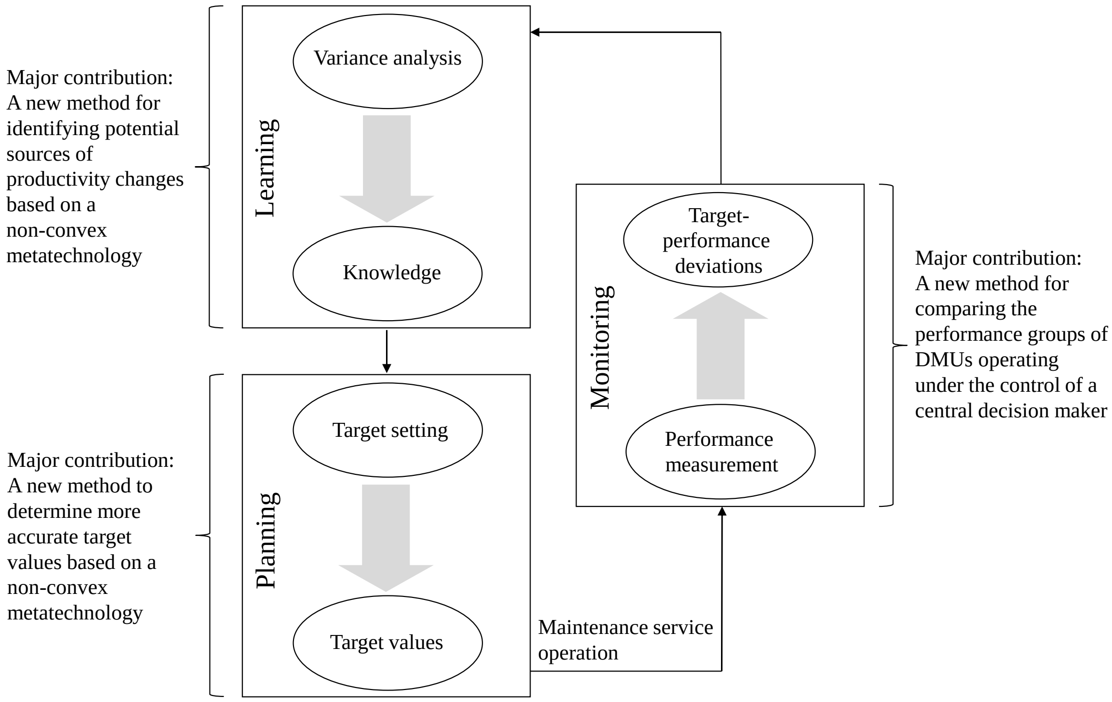
1. *Planning*: The major objective of the planning phase is to provide motivational as well as achievable targets for subordinated operating units. Within this step of the performance management circle, it is tremendously important to apply a realistic set of assumptions about the production opportunities of each DMU. These assumptions define the course of the production frontier and, consequently, the computed target values. An inappropriate estimation of the production frontier may cause too high or too low targets and, therefore, avoids optimal resource usage. In this sense, the introduced approaches of Chapters 5 and 6 provide a more accurate estimation of the metafrontier. As it has been mentioned above, this new type of

global benchmark technology should be applied when convexifications of local technologies are in clear contrast to the underlying organizational setting. In addition to that, the approach in Chapter 6 also accounts for improvement potentials that can be received from centrally coordinated resource reallocations or group-wide collaborations. In sum, both approaches provide more accurate target values than basic DEA models as they explicitly incorporate important features of the organizational settings.

2. *Monitoring*: So far, the relevant literature does not contain a suitable approach for the performance comparison of centrally managed groups. Existing approaches such as the one proposed by Camanho and Dyson (2006) are developed for decentralized management scenarios only. In order to compare distinct sets of operating units, these approaches use an artificial aggregation of the DMU-specific performance scores (e.g., geometric or arithmetic averages). The new approach introduced in Chapter 6 avoids artificial aggregations of single performance scores (e.g., arithmetic or geometric averages) and is capable of capturing directly the performance of the groups on the basis of their internal abilities in transforming inputs to outputs. This is achieved on the basis of the model proposed by Lozano and Villa (2004) (and further extended by Mar-Molinero et al. 2014) which captures the performance of all units simultaneously. Hence, deviations compared to a certain benchmark or target value can be computed with this new approach when centrally managed groups of DMUs are to be evaluated also. Furthermore, the new approach satisfies the circularity property and, therefore, provides new insights when a performance ranking of groups is sought.
3. *Learning*: In the course of continuous efficiency improvement, it may be helpful to track an entity's performance changes over time. This can yield valuable managerial information about the success of strategies or the influence of other business-related factors. In the end, this information may likely help the top management to improve its knowledge about the operating business and, consequently, raise the current performance level of the organization. In this sense, the Malmquist index as proposed in Chapter 5 is designed to measure a unit's performance changes over time. In order to identify potential sources of productivity

changes and, therefore, determine major reasons for target-performance deviations, the index can be decomposed into the two standard Malmquist index components: the efficiency change component and the best practice change component.

Figure 7.1: Major contributions of this thesis[83]



The aforementioned discussions have shown that organizational structures have a substantial influence on a company´s performance and, hence, should be explicitly incorporated in the measurement approach. The study presented here has been designed as a first step to examine how different organizational structures can be mathematically modeled in DEA to improve the overall reliability of the respective performance scores. Recall that the research focus was placed intentionally on the concept of centralization as this variable has a major importance for the case of KONE Corporation. With this in mind, the study presented here can be easily expanded to the examination of other organizational structures (e.g., formalization, configuration) or their combinations. This not only allows a more comprehensive view of the performance of organizations but may also improve the practical applicability of DEA in the near future.

[83] This Figure is based on Ahn (2003, p. 83).

References

Adler, N., L. Friedman, Z. Sinuany-Stern. 2002. Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research*. Vol. 140(2), pp. 249-265.

Adler, N., E. Yazhemsky. 2010. Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *European Journal of Operational Research*. Vol. 202(1), pp. 273-284.

Afsharian, M., H. Ahn. 2014. The Luenberger indicator and directions of measurement: a bottoms-up approach with an empirical illustration to German savings banks. *International Journal of Production Research*. Vol. 52(20), pp. 6216-6233.

Afsharian, M., H. Ahn. 2015. The overall Malmquist index: A new approach for measuring productivity changes over time. *Annals of Operations Research*. Vol. 226(1), pp. 1-27.

Afsharian, M., H. Ahn. 2017. Multi-period productivity measurement under centralized management with an empirical illustration to German saving banks. *OR Spectrum*. Vol. 39(3), pp. 881-911.

Afsharian, M., H. Ahn, S. G. Harms. 2018a. A non-convex meta-frontier Malmquist index for measuring productivity over time. *IMA Journal of Management Mathematics*. Vol. 29(4), pp. 377-392.

Afsharian, M., H. Ahn, S. G. Harms. 2019b. Performance comparison of management groups under centralised management. *European Journal of Operational Research*. Vol. 278(3), pp. 845-854.

Afsharian, M., H. Ahn, A. Lopes, B. Vilela. 2019a. Pitfalls in estimating the X-factor: The case of energy transmission regulation in Brazil. *Socio-Economic Planning Sciences*. Vol. 65(1), pp. 1-9.

Afsharian, M., H. Ahn, L. Neumann. 2016. Generalized DEA: An approach for supporting input/output factor determination in DEA. *Benchmarking – An International Journal*. Vol. 23(7), pp. 1892-1909.

Afsharian, M., H. Ahn, E. Thanassoulis. 2017. A DEA-based incentives system for centrally managed multi-unit organisations. *European Journal of Operational Research*. Vol. 259(2), pp. 587-598.

Afsharian, M., H. Ahn, E. Thanassoulis. 2019c. A frontier-based system of incentives for units in organisations with varying degrees of decentralisation. *European Journal of Operational Research*. Vol. 275(1), pp. 224-237.

Afsharian, M., V. V. Podinovski. 2018. A linear programming approach to efficiency evaluation in nonconvex metatechnologies. *European Journal of Operational Research*. Vol. 268(1), pp. 268-280.

Agrell, P. J., P. Bogetoft. 2010. Endogenous generalized weights under DEA control. *Working paper 2010/02 Louvain School of Management Research Institute.*

Ahn H. 2003. Effektivitäts-und Effizienzsicherung: Controlling-Konzept und Balanced Scorecard. Frankfurt am Main.

Ahn, H. 2014. Data Envelopment Analysis: Mehr als Benchmarking. *Controller Magazin*. Vol. 39(5), pp. 63-65.

Ahn, H., M. Clermont. Performance measurement. In: Corsten H., R. Gössinger, T. Spengler (Eds.), *Handbuch Produktions- und Logistikmanagement in Wertschöpfungsnetzwerken*: Berlin. 2018, pp. 886-903.

Ahn, H., M. Clermont, S. Schwetschke. 2018a. Research on target costing: Past, present and future. *Management Review Quarterly*. Vol. 68(3), pp. 321-354.

Ahn, H., H. Dyckhoff, R. Gilles. 2007. Datenaggregation zur Leistungsbeurteilung durch Ranking: Vergleich der CHE-und DEA-Methodik sowie Ableitung eines Kompromissansatzes. *Journal of Business Economics*. Vol. 77(6), pp. 615-643.

Ahn, H., S. G. Harms, C. Spang. 2018b. Ableitung eines Zielsystems als Basis der Steuerung öffentlicher Theater. *Zeitschrift für öffentliche und gemeinwirtschaftliche Unternehmen*. Vol. 42(3).

Ahn, H., M. H. Le. 2014. An insight into the specification of the input-output set for DEA-based bank efficiency measurement. *Management Review Quarterly*. Vol. 64(1), pp. 3-37.

Ahn, H., M. H. Le. 2015. DEA efficiency of German savings banks – Evidence from a goal-oriented perspective. *Journal of Business Economics*. Vol. 85(9), pp. 953-975.

Ahn, H., M. H. Le. Decision-oriented performance measurement framework. In: Ahn H., M. Clermont, R. Souren (Eds.), *Nachhaltiges Entscheiden – Beiträge zum multiperspektivischen Performancemanagement von Wertschöpfungsprozessen*: Wiesbaden. 2016, pp. 369-383.

Ahn, H., L. Neumann. 2014. Measuring effectiveness – A DEA approach under predetermined targets. *International Journal of Business Analytics*. Vol. 1(1), pp. 16-28.

Ahn, H., L. Neumann, N. Vazquez Novoa. 2012. Measuring the relative balance of DMUs. *European Journal of Operational Research*. Vol. 221(2), pp. 417-423.

Ahn, H., N. Vazquez Novoa. 2018. Balanced data envelopment analysis. *Controlling*. Vol. 30(3), pp. 62-68.

Aigner, D. J., S. F. Chu. 1968. On estimating the industry production function. *The American Economic Review*. Vol. 58(4), pp. 826-839.

Ali, A. I., L. M. Seiford. 1990. Translation invariance in data envelopment analysis. *Operations Research Letters*. Vol. 9(6), pp. 403-405.

Allen, R., A. Athanassopoulos, R. G. Dyson, E. Thanassoulis. 1997. Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research*. Vol. 73, pp. 13-34.

Ameryoun, A., S. Najafi, B. Nejati-Zarnaqi, S. O. Khalilifar, M. Ajam, A. Ansarimoghadam. 2017. Factor selection for service quality evaluation: A hospital case study. *International Journal of Health Care Quality Assurance*. Vol. 30(1), pp. 58-66.

Amin, G. R. 2009. Comments on finding the most efficient DMUs in DEA: An improved integrated model. *Computers & Industrial Engineering*. Vol. 56(4), pp. 1701-1702.

Amin, G. R., M. Toloo. 2007. Finding the most efficient DMUs in DEA: An improved integrated model. *Computers & Industrial Engineering*. Vol. 52(1), pp. 71-77.

Amin, G. R., M. Toloo, B. Sohrabi. 2006. An improved MCDM DEA model for technology selection. *International Journal of Production Research*. Vol. 44(13), pp. 2681-2686.

Amirteimoori, A., A. Emrouznejad. 2011. Input/output deterioration in production processes. *Expert Systems with Applications*. Vol. 38(5), pp. 5822-5825.

Amirteimoori, A., A. Emrouznejad. 2012. Optimal input/output reduction in production processes. *Decision Support Systems*. Vol. 52(3), pp. 742-747.

Amirteimoori, A., S. Kordrostami. 2005. Allocating fixed costs and target setting: A DEA-based approach. *Applied Mathematics and Computation*. Vol. 171(1), pp. 136-151.

Amirteimoori, A., S. Kordrostami. 2013. An alternative clustering approach: A DEA-based procedure. *Optimization*. Vol. 62(2), pp. 227-240.

Amirteimoori, A., S. Kordrostami, A. Masoumzadeh, M. Maghbouli. 2014. Increasing the discrimination power of data envelopment analysis. *International Journal of Operational Research*. Vol. 19(2), pp. 198-210.

Amirteimoori, A., M. Shafiei. 2006. Characterizing an equitable omission of shared resources: A DEA-based approach. *Applied Mathematics and Computation*. Vol. 177(1), pp. 18-23.

Amirteimoori, A., M. M. Tabar. 2010. Resource allocation and target setting in data envelopment analysis. *Expert Systems with Applications*. Vol. 37(4), pp. 3036-3039.

Andersen, P., N. C. Petersen. 1993. A procedure for ranking efficient units in data envelopment analysis. *Management Science*. Vol. 39(10), pp. 1261-1264.

Angulo-Meza, L., M. P. E. Lins. 2002. Review of methods for increasing discrimination in data envelopment analysis. *Annals of Operations Research*. Vol. 116(1-4), pp. 225-242.

Aparicio, J., J. T. P. Ciurana. 2006. New centralized resource allocation DEA models under constant returns to scale. *Working paper I-2009-09 Centro de Investigación Operativa.*

Aparicio, J., E. Crespo-Cebada, F. Pedraja-Chaparro, D. Santín. 2017. Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach. *European Journal of Operational Research*. Vol. 256(2), pp. 533-542.

Aparicio, J., D. Santín. 2018. A note on measuring group performance over time with pseudo-panels. *European Journal of Operational Research*. Vol. 267(1), pp. 227-235.

Armour, H. O., D. J. Teece. 1978. Organizational structure and economic performance: A test of the multidivisional hypothesis. *The Bell Journal of Economics*. Vol. 9(1), pp. 106-122.

Asmild, M., J. C. Paradi, V. Aggarwall, C. Schaffnit. 2004. Combining DEA window analysis with the Malmquist index approach in a study of the Canadian banking industry. *Journal of Productivity Analysis*. Vol. 21(1), pp. 67-89.

Asmild, M., J. C. Paradi, J. T. Pastor. 2009. Centralized resource allocation BCC models. *Omega*. Vol. 37(1), pp. 40-49.

Asosheh, A., S. Nalchigar, M. Jamporazmey. 2010. Information technology project evaluation: An integrated data envelopment analysis and balanced scorecard approach. *Expert Systems with Applications*. Vol. 37(8), pp. 5931-5938.

Athanassopoulos, A. D. 1995. Goal programming & data envelopment analysis (GoDEA) for target-based multi-level planning: Allocating central grants to the Greek local authorities. *European Journal of Operational Research*. Vol. 87(3), pp. 535-550.

Athanassopoulos, A. D. 1998. Decision support for target-based resource allocation of public services in multiunit and multilevel systems. *Management Science*. Vol. 44(2), pp. 173-187.

Atkinson, A. A., J. H. Waterhouse, R. B. Wells. 1997. A stakeholder approach to strategic performance measurement. *MIT Sloan Management Review*. Vol. 38(3), pp. 25-37.

Azar, A., M. Zarei Mahmoudabadi, A. Emrouznejad. 2016. A new fuzzy additive model for determining the common set of weights in data envelopment analysis. *Journal of Intelligent & Fuzzy Systems*. Vol. 30(1), pp. 61-69.

Azizi, H. 2011. The interval efficiency based on the optimistic and pessimistic points of view. *Applied Mathematical Modelling*. Vol. 35(5), pp. 2384-2393.

Babalos, V., M. Doumpos, N. Philippas, C. Zopounidis. 2015. Towards a holistic approach for mutual fund performance appraisal. *Computational Economics*. Vol. 46(1), pp. 35-53.

Balk, B. M. 2001. Scale efficiency and productivity change. *Journal of Productivity Analysis*. Vol. 15(3), pp. 159-183.

Banker, R. D., A. Charnes, W. W. Cooper. 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*. Vol. 30(9), pp. 1078-1092.

Banker, R. D., W. W. Cooper, L. M. Seiford, R. M. Thrall, J. Zhu. 2004. Returns to scale in different DEA models. *European Journal of Operational Research*. Vol. 154(2), pp. 345-362.

Barney, J. B., P. M. Wright. 1998. On becoming a strategic partner: The role of human resources in gaining competitive advantage. *Working paper 4-1-1997 Cornell University ILR School.*

Barr, R. S. DEA software tools and technology: A state-of-the-art survey. In: Cooper W. W., L. M. Seiford, J. Zhu (Eds.), *Handbook on data envelopment analysis*: Boston, MA. 2004, pp. 539-566.

Battese, G. E., D. S. P. Rao. 2002. Technology gap, efficiency, and a stochastic metafrontier function. *International Journal of Business and Economics*. Vol. 1(2), pp. 87-93.

Battese, G. E., D. S. P. Rao, C. J. O'Donnell. 2004. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*. Vol. 21(1), pp. 91-103.

Bea F. X., E. Göbel. 2010. Organisation: Theorie und Gestaltung. Stuttgart.

Beasley, J. E. 2003. Allocating fixed costs and resources via data envelopment analysis. *European Journal of Operational Research*. Vol. 147(1), pp. 198-216.

Berg, S. A., F. R. Førsund, E. S. Jansen. 1992. Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *The Scandinavian Journal of Economics*. Vol. 94, pp. 211-228.

Berger, A. N., D. B. Humphrey. 1997. Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research*. Vol. 98(2), pp. 175-212.

Bi, G., J. Ding, Y. Luo, L. Liang. 2011. Resource allocation and target setting for parallel production system based on DEA. *Applied Mathematical Modelling*. Vol. 35(9), pp. 4270-4280.

Bible, L., S. Kerr, M. Zanini. 2006. The balanced scorecard – Here and back: From its beginnings as a performance measurement tool. *Management Accounting Quarterly*. Vol. 7(4), pp. 18-23.

Blancas, F. J., I. Contreras, J. M. Ramírez-Hurtado. 2013. Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives. *Journal of the Operational Research Society*. Vol. 64(5), pp. 668-678.

Blenkinsop, S., L. Davis. 1991. The road to continuous improvement. *OR Insight*. Vol. 4(3), pp. 23-26.

Borzoei, S., M. Zohrehbandian. 2013. Common weights for the evaluation of decision-making units with nonlinear virtual inputs and outputs. *International Journal of Data Envelopment Analysis*. Vol. 1(3), pp. 167-173.

Bouyssou, D. 1999. Using DEA as a tool for MCDM: Some remarks. *Journal of the Operational Research Society*. Vol. 50(9), pp. 974-978.

Bowlin, W. F. 1987. Evaluating the efficiency of US air force real-property maintenance activities. *Journal of the Operational Research Society*. Vol. 38(2), pp. 127-135.

Breustedt, G., T. Francksen, U. Latacz-Lohmann. Estimating non-concave metafrontiers using data envelope analysis. In: Glebe T., A. Heißenhuber, L. Kirner, S. Pöchtrager (Eds.), *Agrar- und Ernährungswirtschaft im Umbruch: Schriften der Gesellschaft für Wirtschafts- und Sozialwissenschaften des Landbaues e.V.*, Vol. 43: Münster-Hiltrup. 2008, pp. 531-542.

Burgess, J. F., P. W. Wilson. 1995. Decomposing hospital productivity changes, 1985-1988: A nonparametric Malmquist approach. *Journal of Productivity Analysis*. Vol. 6(4), pp. 343-363.

Burton, R. M., B. Obel. The Environment. In: *Strategic Organizational Diagnosis and Design: The Dynamics of Fit*: Boston. 2004, pp. 185-235.

Camanho, A. S., R. G. Dyson. 2006. Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of Productivity Analysis*. Vol. 26(1), pp. 35-49.

Caves, D. W., L. R. Christensen, W. E. Diewert. 1982. The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*. Vol. 50(6), pp. 1393-1414.

Chambers, R. G., R. Färe, S. Grosskopf. 1996. Productivity growth in APEC countries. *Pacific Economic Review*. Vol. 1(3), pp. 181-190.

Charnes, A., C. T. Clark, W. W. Cooper, B. Golany. 1984. A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces. *Annals of Operations Research*. Vol. 2(1), pp. 95-112.

Charnes, A., W. W. Cooper. 1962. Programming with linear fractional functionals. *Naval Research Logistics Quarterly*. Vol. 9(3-4), pp. 181-186.

Charnes, A., W. W. Cooper, B. Golany, L. Seiford, J. Stutz. 1985. Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of econometrics*. Vol. 30(1-2), pp. 91-107.

Charnes A., W. W. Cooper, A. Y. Lewin, L. M. Seiford. 2013. Data envelopment analysis: Theory, methodology, and applications. New York.

Charnes, A., W. W. Cooper, E. Rhodes. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*. Vol. 2(6), pp. 429-444.

Charnes, A., W. W. Cooper, E. Rhodes. 1979. Short communication: Measuring the efficiency of decision making units. *European Journal of Operational Research*. Vol. 3(4), pp. 339.

Charnes, A., W. W. Cooper, E. Rhodes. 1981. Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*. Vol. 27(6), pp. 668-697.

Chen, J.-X. 2011. Peer-estimation for multiple criteria ABC inventory classification. *Computers & Operations Research*. Vol. 38(12), pp. 1784-1791.

Chen, K.-H., H.-Y. Yang. 2011. A cross-country comparison of productivity growth using the generalised metafrontier Malmquist productivity index: With application to banking industries in Taiwan and China. *Journal of Productivity Analysis*. Vol. 35(3), pp. 197-212.

Chen, L., F.-M. Wu, F. Feng, F. Lai, Y.-M. Wang. 2018. A common set of weights for ranking decision-making units with undesirable outputs: A double frontiers data envelopment analysis approach. *Asia-Pacific Journal of Operational Research*. Vol. 35(6), pp. 1-25.

Chen, Y.-W., M. Larbani, Y.-P. Chang. 2009. Multiobjective data envelopment analysis. *Journal of the Operational Research Society*. Vol. 60(11), pp. 1556-1566.

Choi, Y., D.-H. Oh, N. Zhang. 2015. Environmentally sensitive productivity growth and its decompositions in China: A metafrontier Malmquist-Luenberger productivity index approach. *Empirical Economics*. Vol. 49(3), pp. 1017-1043.

Clarke, R. L. 1992. Evaluating USAF vehicle maintenance productivity over time: An application of data envelopment analysis. *Decision Sciences*. Vol. 23(2), pp. 376-384.

Coelli, T. J., D. S. P. Rao. 2005. Total factor productivity growth in agriculture: A Malmquist index analysis of 93 countries, 1980-2000. *Agricultural Economics*. Vol. 32(1), pp. 115-134.

Coelli T. J., D. S. P. Rao, C. J. O'Donnell, G. E. Battese. 2005. An introduction to efficiency and productivity analysis.

Cook, W. D., D. Chai, J. Doyle, R. Green. 1998. Hierarchies and groups in DEA. *Journal of Productivity Analysis*. Vol. 10(2), pp. 177-198.

Cook, W. D., M. Hababou, H. J. H. Tuenter. 2000. Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches. *Journal of Productivity Analysis*. Vol. 14(3), pp. 209-224.

Cook, W. D., A. Kazakov, Y. Roll. On the measurement and monitoring of relative efficiency of highway maintenance patrols. In: Charnes A., W. W. Cooper, A. Y. Lewin, L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology, and applications*: Dordrecht. 1994, pp. 195-210.

Cook, W. D., A. Kazakov, Y. Roll, L. M. Seiford. 1991. A data envelopment approach to measuring efficiency: Case analysis of highway maintenance patrols. *The Journal of Socio-Economics*. Vol. 20(1), pp. 83-103.

Cook, W. D., M. Kress. 1990. A data envelopment model for aggregating preference rankings. *Management Science*. Vol. 36(11), pp. 1302-1310.

Cook, W. D., M. Kress, L. M. Seiford. 1996. Data envelopment analysis in the presence of both quantitative and qualitative factors. *Journal of the Operational Research Society*. Vol. 47(7), pp. 945-953.

Cook, W. D., J. L. Ruiz, I. Sirvent, J. Zhu. 2017. Within-group common benchmarking using DEA. *European Journal of Operational Research*. Vol. 256(3), pp. 901-910.

Cook, W. D., J. Zhu. 2007. Within-group common weights in DEA: An analysis of power plant efficiency. *European Journal of Operational Research*. Vol. 178(1), pp. 207-216.

Cooper, W. W., J. L. Ruiz, I. Sirvent. Choices and uses of DEA weights. In: Cooper W. W., L. M. Seiford, J. Zhu (Eds.), *Handbook on data envelopment analysis*: Boston. 2011a, pp. 93-126.

Cooper W. W., L. M. Seiford, K. Tone. 2006. Introduction to data envelopment analysis and its uses: With DEA-solver software and references. New York.

Cooper W. W., L. M. Seiford, K. Tone. 2007. Data envelopment analysis: A comprehensive text with models, applications, references and DEA-solver software. New York.

Cooper, W. W., L. M. Seiford, J. Zhu. Data envelopment analysis: History, models, and interpretations. In: Cooper W. W., L. M. Seiford, J. Zhu (Eds.), *Handbook on data envelopment analysis*: Boston. 2011b, pp. 1-39.

Cordes-Berszinn P. 2013. Dynamic capabilities: How organizational structures affect knowledge processes. London.

Crawford, K. M., J. F. Cox. 1990. Designing performance measurement systems for just-in-time operations. *International Journal of Production Research*. Vol. 28(11), pp. 2025-2036.

Cray, D. 1984. Control and coordination in multinational corporations. *Journal of International Business Studies*. Vol. 15(2), pp. 85-98.

Cross, K. F., R. L. Lynch. 1988. The "SMART" way to define and sustain success. *National Productivity Review*. Vol. 8(1), pp. 23-33.

Daft R. L., J. Murphy, H. Willmott. 2010. Organization theory and design. Boston.

Davoodi, A., H. Z. Rezai. 2012. Common set of weights in data envelopment analysis: A linear programming problem. *Central European Journal of Operations Research*. Vol. 20(2), pp. 355-365.

Demski J. 1997. Managerial uses of accounting information. Norwell.

Dent, J. F. 1990. Strategy, organization and control: Some possibilities for accounting research. *Accounting, Organizations and Society*. Vol. 15(1-2), pp. 3-25.

Despotis, D. K. 2002. Improving the discriminating power of DEA: Focus on globally efficient units. *Journal of the Operational Research Society*. Vol. 53(3), pp. 314-323.

Despotis, D. K. 2005. Measuring human development via data envelopment analysis: The case of Asia and the Pacific. *Omega*. Vol. 33(5), pp. 385-390.

Dispan, J. 2007. Aufzüge und Fahrtreppen: Eine Branche im Wandel. *Lift Report*. Vol. 33(6), pp. 24-38.

Dispan, J. 2015. Aufzugs- und Fahrtreppenbranche in Deutschland: Entwicklungstrends und Herausforderungen – Branchenreport 2015. *Informationsdienst des IMU Instituts*. Vol. 2015(1), pp. 1-83.

Dixon J. R. 1990. The new performance challenge: Measuring operations for world-class competition. Homewood.

Domínguez-Serrano, M., F. J. Blancas. 2011. A gender wellbeing composite indicator: The best-worst global evaluation approach. *Social Indicators Research*. Vol. 102(3), pp. 477-496.

Doyle, J. R. 1995. Multiattribute choice for the lazy decision maker: Let the alternatives decide! *Organizational Behavior and Human Decision Processes*. Vol. 62(1), pp. 87-100.

Du, J., W. D. Cook, L. Liang, J. Zhu. 2014. Fixed cost and resource allocation based on DEA cross-efficiency. *European Journal of Operational Research*. Vol. 235(1), pp. 206-214.

Du, J., L. Liang, Y. Chen, G.-b. Bi. 2010. DEA-based production planning. *Omega*. Vol. 38(1-2), pp. 105-112.

Dyckhoff, H. 1985. Kompensation bei Entscheidungskriterien. *OR Spectrum*. Vol. 7(4), pp. 195-207.

Dyckhoff, H. Analyse der Werbeeffizienz mit Advanced DEA. In: Bauer H., M. Staat, M. Hammerschmidt (Eds.), *Marketingeffizienz-Messung und Steuerung mit der DEA: Konzept und Einsatz in der Praxis*: München. 2006, pp. 167-186.

Dyckhoff, H., H. Ahn. 2010. Verallgemeinerte DEA-Modelle zur Performanceanalyse. *Zeitschrift für Betriebswirtschaft*. Vol. 80(12), pp. 1249-1276.

Dyckhoff, H., M. Clermont, A. Dirksen, E. Mbock. Measuring balanced effectiveness and efficiency of German business schools' research performance. In: Dilger A., H. Dyckhoff, G. Fandel (Eds.), *Performance Management im Hochschulbereich. ZfB-Special Issues.*, Vol. 3: Wiesbaden. 2013, pp. 39-60.

Dyckhoff, H., S. Gutgesell. 2015. Properties of DEA-integrated balance and specialization measures. *OR Spectrum*. Vol. 37(2), pp. 503-527.

Dyson, R. G., R. Allen, A. S. Camanho, V. V. Podinovski, C. S. Sarrico, E. A. Shale. 2001. Pitfalls and protocols in DEA. *European Journal of Operational Research*. Vol. 132(2), pp. 245-259.

Dyson, R. G., E. Thanassoulis. 1988. Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*. Vol. 39(6), pp. 563-576.

Ebrahimnejad, A. 2012. A new approach for ranking of candidates in voting systems. *OPSEARCH*. Vol. 49(2), pp. 103-115.

Emrouznejad, A., B. R. Parker, G. Tavares. 2008. Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences*. Vol. 42(3), pp. 151-157.

Emrouznejad, A., G.-l. Yang. 2018. A survey and analysis of the first 40 years of scholarly literature in DEA: 1978-2016. *Socio-Economic Planning Sciences*. Vol. 61, pp. 4-8.

Ertay, T., D. Ruan, U. R. Tuzkaya. 2006. Integrating data envelopment analysis and analytic hierarchy for the facility layout design in manufacturing systems. *Information Sciences*. Vol. 176(3), pp. 237-262.

Ewert R., A. Wagenhofer. 2014. Interne Unternehmensrechnung. Berlin.

Fallah-Fini, S., K. Triantis, J. M. de la Garza, W. L. Seaver. 2012. Measuring the efficiency of highway maintenance contracting strategies: A bootstrapped non-parametric meta-frontier approach. *European Journal of Operational Research*. Vol. 219(1), pp. 134-145.

Fang, L., C. Q. Zhang. 2008. Resource allocation based on the DEA model. *Journal of the Operational Research Society*. Vol. 59(8), pp. 1136-1141.

Färe, R., R. Grabowski, S. Grosskopf, S. Kraft. 1997a. Efficiency of a fixed but allocatable input: A non-parametric approach. *Economics Letters*. Vol. 56(2), pp. 187-193.

Färe, R., S. Grosskopf, K. Kerstens, J. E. Kirkley, D. Squires. Assessing short-run and medium-run fishing capacity at the industry level and its reallocation. In: Johnston R. S., A. L. Shriver (Eds.), *Microbehavior and macroresults: Proceedings of the tenth biennial conference of the International Institute of Fisheries Economics and Trade*: Corvallis. 2000, pp. 1-10.

Färe, R., S. Grosskopf, B. Lindgren, P. Roos. 1992a. Productivity changes in Swedish pharamacies 1980-1989: A non-parametric Malmquist approach. *Journal of Productivity Analysis*. Vol. 3(1-2), pp. 85-101.

Färe, R., S. Grosskopf, B. Lindgren, P. Roos. Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach. In: *International Applications of Productivity and Efficiency Analysis*. 1992b, pp. 81-97.

Färe, R., S. Grosskopf, B. Lindgren, P. Roos. Productivity developments in Swedish hospitals: A Malmquist output index approach. In: Charnes A., W. W. Cooper, A. Y. Lewin, L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology, and applications*: Dordrecht. 1994, pp. 253-272.

Färe, R., S. Grosskopf, M. Norris. 1997b. Productivity growth, technical progress, and efficiency change in industrialized countries: Reply. *The American Economic Review*. Vol. 87(5), pp. 1040-1044.

Farrell, M. J. 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*. Vol. 120(3), pp. 253-281.

Faust M., P. Jauch, K. Brünnecke, C. Deutschmann. 1994. Dezentralisierung von Unternehmen: Bürokratie-und Hierarchieabbau und die Rolle betrieblicher Arbeitspolitik. München.

Fethi, M. D., F. Pasiouras. 2010. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research*. Vol. 204(2), pp. 189-198.

Foroughi, A. A. 2011a. A new mixed integer linear model for selecting the best decision making units in data envelopment analysis. *Computers & Industrial Engineering*. Vol. 60(4), pp. 550-554.

Foroughi, A. A. 2011b. A note on "A new method for ranking discovered rules from data mining by DEA", and a full ranking approach. *Expert Systems with Applications*. Vol. 38(10), pp. 12913-12916.

Foroughi, A. A. 2012. A modified common weight model for maximum discrimination in technology selection. *International Journal of Production Research*. Vol. 50(14), pp. 3841-3846.

Foroughi, A. A., B. Aouni. 2012. New approaches for determining a common set of weights for a voting system. *International Transactions in Operational Research*. Vol. 19(4), pp. 521-530.

Foroughi, A. A., M. Tamiz. 2005. An effective total ranking model for a ranked voting system. *Omega*. Vol. 33(6), pp. 491-496.

Fraser, I., D. Cordina. 1999. An application of data envelopment analysis to irrigated dairy farms in Northern Victoria, Australia. *Agricultural Systems*. Vol. 59(3), pp. 267-282.

Fredrickson, J. W. 1986. The strategic decision process and organizational structure. *Academy of Management Review*. Vol. 11(2), pp. 280-297.

Friedman, L., Z. Sinuany-Stern. 1997. Scaling units via the canonical correlation analysis in the DEA context. *European Journal of Operational Research*. Vol. 100(3), pp. 629-637.

Garg, A., S. G. Deshmukh. 2006. Maintenance management: Literature review and directions. *Journal of Quality in Maintenance Engineering*. Vol. 12(3), pp. 205-238.

Gattoufi, S., M. Oral, A. Reisman. 2004. Data envelopment analysis literature: A bibliography update (1951-2001). *Socio-Economic Planning Sciences*. Vol. 38(2-3), pp. 159-229.

Ghalayini, A. M., J. S. Noble. 1996. The changing basis of performance measurement. *International Journal of Operations & Production Management*. Vol. 16(8), pp. 63-80.

Gharakhani, D., A. T. Eshlaghy, K. F. Hafshejani, R. K. Mavi, F. H. Lotfi. 2018. Common weights in dynamic network DEA with goal programming approach for performance assessment of insurance companies in Iran. *Management Research Review*. Vol. 41(8), pp. 920-938.

Gilles, R. Performance Measurement mittels Data Envelopment Analysis: Theoretisches Grundkonzept und universitäre Forschungsperformance als Anwendungsfall. In: Müller-Böling D., R. Schulte (Eds.), *Wissenschafts- und Hochschulmanagement*: Köln. 2005.

Globerson, S. 1985. Issues in developing a performance criteria system for an organization. *International Journal of Production Research*. Vol. 23(4), pp. 639-646.

Golany, B., F. Y. Phillips, J. J. Rousseau. 1993. Models for improved effectiveness based on DEA efficiency results. *IIE transactions*. Vol. 25(6), pp. 2-10.

Golany, B., Y. Roll. 1989. An application procedure for DEA. *Omega*. Vol. 17(3), pp. 237-250.

Golany, B., E. Tamir. 1995. Evaluating efficiency-effectiveness-equality trade-offs: A data envelopment analysis approach. *Management Science*. Vol. 41(7), pp. 1172-1184.

Golany, B., G. Yu. 1995. A goal programming-discriminant function approach to the estimation of an empirical production function based on DEA results. *Journal of Productivity Analysis*. Vol. 6(2), pp. 171-186.

Graubner M. 2006. Task, firm size, and organizational structure in management consulting: An empirical analysis from a contingency perspective. Wiesbaden.

Grifell-Tatjé, E., C. A. K. Lovell. 1995. A note on the Malmquist productivity index. *Economics Letters*. Vol. 47(2), pp. 169-175.

Groote, P. de. 1995. Maintenance performance analysis: A practical approach. *Journal of Quality in Maintenance Engineering*. Vol. 1(2), pp. 4-24.

Guedes de Avellar, J. V., A. Z. Milioni, T. N. Rabello. 2007. Spherical frontier DEA model based on a constant sum of inputs. *Journal of the Operational Research Society*. Vol. 58(9), pp. 1246-1251.

Hackman S. 2008. Production economics: Integrating the microeconomic and engineering perspectives. Heidelberg.

Hammann, P. 1976. Choice of the organization structure: A framework for quantitative analysis of industrial centralization/decentralization issues. *Zeitschrift für Operations Research*. Vol. 20(2), B17-B35.

Hao, Q., H. Kasper, J. Muehlbacher. 2012. How does organizational structure influence performance through learning and innovation in Austria and China. *Chinese Management Studies*. Vol. 6(1), pp. 36-52.

Harms S. G. 2016. Performance benchmarking of maintenance groups in the elevator and escalator Industry (Master thesis). Braunschweig.

Hashimoto, A., D.-A. Wu. 2004. A DEA-compromise programming model for comprehensive ranking. *Journal of the Operations Research Society of Japan*. Vol. 47(2), pp. 73-81.

Hatami-Marbini, A., M. Tavana, P. J. Agrell, F. H. Lotfi, Z. G. Beigi. 2015. A common-weights DEA model for centralized resource reduction and target setting. *Computers & Industrial Engineering*. Vol. 79, pp. 195-203.

Hatefi, S. M., S. A. Torabi. 2010. A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics*. Vol. 70(1), pp. 114-120.

Hatefi, S. M., S. A. Torabi. 2015. A common weight linear optimization approach for multicriteria ABC inventory classification. *Advances in Decision Sciences*. Vol. 2015.

Hayami, Y. 1969. Sources of agricultural productivity gap among selected countries. *American Journal of Agricultural Economics*. Vol. 51(3), pp. 564-575.

Hayami, Y., V. W. Ruttan. 1970. Agricultural productivity differences among countries. *The American Economic Review*. Vol. 60(5), pp. 895-911.

Hjalmarsson, L., J. Odeck. 1996. Efficiency of trucks in road construction and maintenance: An evaluation with data envelopment analysis. *Computers & Operations Research*. Vol. 23(4), pp. 393-404.

Horngren C., G. Sundem, W. Stratton. 1996. Introduction to management accouting. New Jersey.

Howell, R. A., S. R. Soucy. 1987. Operating controls in the new manufacturing environment (Part 3). *Strategic Finance*. Vol. 69(4), pp. 25.

Huang, C. W., C. T. Ting, C. H. Lin, C. T. Lin. 2013. Measuring non-convex metafrontier efficiency in international tourist hotels. *Journal of the Operational Research Society*. Vol. 64(2), pp. 250-259.

Ittner, C. D., D. F. Larcker. 1998. Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*. Vol. 36, pp. 1-35.

Ittner, C. D., D. F. Larcker, T. Randall. 2003. Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society*. Vol. 28(7-8), pp. 715-741.

Jacobs, R. 2001. Alternative methods to examine hospital efficiency: Data envelopment analysis and stochastic frontier analysis. *Health Care Management Science*. Vol. 4(2), pp. 103-115.

Jahanshahloo, G. R., F. H. Lotfi, M. Khanmohammadi, M. Kazemimanesh, V. Rezaie. 2010. Ranking of units by positive ideal DMU with common weights. *Expert Systems with Applications*. Vol. 37(12), pp. 7483-7488.

Jahanshahloo, G. R., A. Memariani, F. H. Lotfi, H. Z. Rezai. 2005. A note on some of DEA models and finding efficiency and complete ranking using common set of weights. *Applied Mathematics and Computation*. Vol. 166(2), pp. 265-281.

Jahanshahloo, G. R., J. Sadeghi, M. Khodabakhshi. 2017. Proposing a method for fixed cost allocation using DEA based on the efficiency invariance and common set of weights principles. *Mathematical Methods of Operations Research*. Vol. 85(2), pp. 223-240.

Jahanshahloo, G. R., M. Zohrehbandian, A. Alinezhad, S. A. Naghneh, H. Abbasian, R. K. Mavi. 2011. Finding common weights based on the DM's preference information. *Journal of the Operational Research Society*. Vol. 62(10), pp. 1796-1800.

Joro T., P. J. Korhonen. 2015. Extension of data envelopment analysis with preference information: Value efficiency. Heidelberg.

Kao, C. 2010. Malmquist productivity index based on common-weights DEA: The case of Taiwan forests after reorganization. *Omega*. Vol. 38(6), pp. 484-491.

Kao, C., H.-T. Hung. 2005. Data envelopment analysis with common weights: The compromise solution approach. *Journal of the Operational Research Society*. Vol. 56(10), pp. 1196-1203.

Kaplan, R. S. 2001. Strategic performance measurement and management in nonprofit organizations. *Nonprofit management and Leadership*. Vol. 11(3), pp. 353-370.

Kaplan R. S., D. P. Norton. 1996a. The balanced scorecard: Translating strategy into action. Boston.

Kaplan R. S., D. P. Norton. 1996b. Using the balanced scorecard as a strategic management system. Boston.

Karsak, E. E., S. S. Ahiska. 2005. Practical common weight multi-criteria decision-making approach with an improved discriminating power for technology selection. *International Journal of Production Research*. Vol. 43(8), pp. 1537-1554.

Keegan, D. P., R. G. Eiler, C. R. Jones. 1989. Are your performance measures obsolete? *Strategic Finance*. Vol. 70(12), pp. 45.

Kerstens, K., C. O'Donnell, I. van de Woestyne. 2019. Metatechnology frontier and convexity: A restatement. *European Journal of Operational Research*. Vol. 275(2), pp. 780-792.

Khalili, M., A. S. Camanho, M. C. A. S. Portela, M. R. Alirezaee. 2010. The measurement of relative efficiency using data envelopment analysis with assurance regions that link inputs and outputs. *European Journal of Operational Research*. Vol. 203(3), pp. 761-770.

Khurana, S., R. Banerjee, U. Gaitonde. 2002. Energy balance and cogeneration for a cement plant. *Applied Thermal Engineering*. Vol. 22(5), pp. 485-494.

Kiani Mavi, R., S. Kazemi, J. M. Jahangiri. 2013. Developing common set of weights with considering nondiscretionary inputs and using ideal point method. *Journal of Applied Mathematics*. Vol. 2013.

Kiani Mavi, R., A. Makui, S. Fazli, A. Alinezhad. 2010. Practical common weights compromise solution approach for technology selection. *International Journal of Procurement Management*. Vol. 3(2), pp. 214-230.

Kieser A., P. Walgenbach. 2010. Organisation. Stuttgart.

Kleine, A., B. Glaser. 2004. Distance maximization in DEA. *Working paper, University of Hohenheim.*

Kogelheide B. 1992. Entwicklung realer Organisationsstrukturen: Eine lebenszyklus-orientierte Analyse. Wiesbaden.

KONE. 2017a. President and CEO of KONE to change as of April 1, 2014. Accessed on 13 Apr 2017. http://www.kone.co.uk/about-us/news-events/news/president-and-ceo-of-kone-to-change-as-of-april-1-2014-2014-01-28.aspx.

KONE. 2017b. Verschärfte Vorschriften für Aufzugbetreiber. Accessed on 13 Apr 2017. http://www.kone.de/unternehmen/presse-tv-events/news/Informationen-zur-Betriebssicherheitsverordnung.aspx.

KONE. 2018a. Key figures. Accessed on 9 Nov 2018. https://www.kone.com/en/investors/financials-and-outlook/key-figures/.

KONE. 2018b. KONE in brief. Accessed on 9 Nov 2018. https://www.kone.com/en/Images/kone-in-brief-2017_tcm17-43142.pdf.

KONE. 2019. Key figures. Accessed on 3 Mar 2019. https://www.kone.com/en/investors/financials-and-outlook/key-figures/.

Kontolaimou, A., K. Tsekouras. 2010. Are cooperatives the weakest link in European banking? A non-parametric metafrontier approach. *Journal of banking & Finance*. Vol. 34(8), pp. 1946-1957.

Koontz H., H. Weihrich. 2010. Essentials of management: An international perspective. New Delhi.

Krasman, J. 2011. Taking feedback-seeking to the next" level": Organizational structure and feedback-seeking behavior. *Journal of Managerial Issues*. Vol. 23(1), pp. 9-30.

Kutucuoglu, K. Y., J. Hamali, Z. Irani, J. M. Sharp. 2001. A framework for managing maintenance using performance measurement systems. *International Journal of Operations & Production Management*. Vol. 21(1-2), pp. 173-195.

Latruffe, L., J. Fogarasi, Y. Desjeux. 2012. Efficiency, productivity and technology comparison for farms in Central and Western Europe: The case of field crop and dairy farming in Hungary and France. *Economic Systems*. Vol. 36(2), pp. 264-278.

Le M. H. 2015. A decision-oriented approach to performance measurement. Frankfurt am Main et al.

Lebas, M. J. 1995. Performance measurement and performance management. *International Journal of Production Economics*. Vol. 41(1-3), pp. 23-35.

Lee, J.-D., C. Baek, H.-S. Kim, J.-S. Lee. 2014. Development pattern of the DEA research field: A social network analysis approach. *Journal of Productivity Analysis*. Vol. 41(2), pp. 175-186.

Li, F., J. Song, A. Dolgui, L. Liang. 2017. Using common weights and efficiency invariance principles for resource allocation and target setting. *International Journal of Production Research*. Vol. 55(17), pp. 4982-4997.

Li, F., Q. Zhu, L. Liang. 2018. Allocating a fixed cost based on a DEA-game cross efficiency approach. *Expert Systems with Applications*. Vol. 96, pp. 196-207.

Li, Y., F. Yang, L. Liang, Z. Hua. 2009. Allocating the fixed cost as a complement of other cost inputs: A DEA approach. *European Journal of Operational Research*. Vol. 197(1), pp. 389-401.

Liu, F.-H. F., H. H. Peng. 2008. Ranking of units on the DEA frontier with common weights. *Computers & Operations Research*. Vol. 35(5), pp. 1624-1637.

Liu, F.-H. F., H. H. Peng. 2009. A systematic procedure to obtain a preferable and robust ranking of units. *Computers & Operations Research*. Vol. 36(4), pp. 1012-1025.

Lotfi, F. H., A. Hatami-Marbini, P. J. Agrell, N. Aghayi, K. Gholami. 2013. Allocating fixed resources and setting targets using a common-weights DEA approach. *Computers & Industrial Engineering*. Vol. 64(2), pp. 631-640.

Lotfi, F. H., G. R. Jahanshahloo, A. Memariani. 2000. A method for finding common set of weights by multiple objective programming in data envelopment analysis. *Southwest Journal of Pure and Applied Mathematics*. Vol. 2000(1), pp. 44-54.

Lovell, C. A. K. 2003. The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis*. Vol. 20(3), pp. 437-458.

Lozano, S. 2014. Nonradial approach to allocating fixed-costs and common revenue using centralized DEA. *International Journal of Information Technology & Decision Making*. Vol. 13(1), pp. 29-46.

Lozano, S., G. Villa. 2004. Centralized resource allocation using data envelopment analysis. *Journal of Productivity Analysis*. Vol. 22(1-2), pp. 143-161.

Lozano, S., G. Villa. 2005. Centralized DEA models with the possibility of downsizing. *Journal of the Operational Research Society*. Vol. 56(4), pp. 357-364.

Lozano, S., G. Villa, B. Adenso-Díaz. 2004. Centralised target setting for regional recycling operations using DEA. *Omega*. Vol. 32(2), pp. 101-110.

Lozano, S., G. Villa, D. Canca. 2011. Application of centralised DEA approach to capital budgeting in Spanish ports. *Computers & Industrial Engineering*. Vol. 60(3), pp. 455-465.

Makui, A., A. Alinezhad, R. Kiani Mavi, M. Zohrehbandian. 2008. A goal programming method for finding common weights in DEA with an improved discriminating power for efficiency. *Journal of Industrial and Systems Engineering*. Vol. 1(4), pp. 293-303.

Makui, A., M. Momeni. 2012. Using CSW weight's in UTASTAR method. *Decision Science Letters*. Vol. 1(1), pp. 39-46.

Malmquist, S. 1953. Index numbers and indifference surfaces. *Trabajos de Estadistica y de Investigacion Operativa*. Vol. 4(2), pp. 209-242.

Mar-Molinero, C., D. Prior, M.-M. Segovia, F. Portillo. 2014. On centralized resource utilization and its reallocation by using DEA. *Annals of Operations Research*. Vol. 221(1), pp. 273-283.

Maskell B. H. 1991. Performance measurement for world class manufacturing: A model for American companies. New York.

Medal-Bartual, A., C.-J. Garcia-Martin, R. Sala-Garrido. 2012. Efficiency analysis of small franchise enterprises through a DEA metafrontier model. *The Service Industries Journal*. Vol. 32(15), pp. 2421-2434.

Miah A. Q. 2016. Applied statistics for social and management sciences. Singapore.

Mintzberg H. 1979. The structuring of organization: A synthesis of the research. Englewood Cliffs.

Muchiri, P., L. Pintelon, L. Gelders, H. Martin. 2011. Development of maintenance function performance measurement framework and indicators. *International Journal of Production Economics*. Vol. 131(1), pp. 295-302.

Nakabayashi, K., K. Tone. 2006. Egoist's dilemma: A DEA game. *Omega*. Vol. 34(2), pp. 135-148.

Nardo M., M. Saisana, A. Saltelli, S. Tarantola, A. Hoffman, E. Giovannini. 2005. Handbook on constructing composite indicators: Methodology and user guide. Paris.

Neely, A., M. Gregory, K. Platts. 1995. Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management*. Vol. 15(4), pp. 80-116.

Nesterenko, V., V. Zelenyuk. 2007. Measuring potential gains from reallocation of resources. *Journal of Productivity Analysis*. Vol. 28(1-2), pp. 107-116.

O'Donnell, C. J., D. S. P. Rao, G. E. Battese. 2008. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*. Vol. 34(2), pp. 231-255.

Oh, D.-H. 2010. A global Malmquist-Luenberger productivity index. *Journal of Productivity Analysis*. Vol. 34(3), pp. 183-197.

Oh, D.-H., J.-D. Lee. 2010. A metafrontier approach for measuring Malmquist productivity index. *Empirical Economics*. Vol. 38(1), pp. 47-64.

Olson, E. M., S. F. Slater, G. T. M. Hult. 2005. The performance implications of fit among business strategy, marketing organization structure, and strategic behavior. *Journal of Marketing*. Vol. 69(3), pp. 49-65.

Omrani, H. 2013. Common weights data envelopment analysis with uncertain data: A robust optimization approach. *Computers & Industrial Engineering*. Vol. 66(4), pp. 1163-1170.

Ozcan Y. A. 2008. Health care benchmarking and performance evaluation: An assessment using data envelopment analysis (DEA). Richmond.

Pachkova, E. V. 2009. Restricted reallocation of resources. *European Journal of Operational Research*. Vol. 196(3), pp. 1049-1057.

Paradi, J. C., S. Rouatt, H. Zhu. 2011. Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega*. Vol. 39(1), pp. 99-109.

Parida, A., G. Chattopadhyay. 2007. Development of a multi-criteria hierarchical framework for maintenance performance measurement (MPM). *Journal of Quality in Maintenance Engineering*. Vol. 13(3), pp. 241-258.

Parida, A., U. Kumar. 2006. Maintenance performance measurement (MPM): Issues and challenges. *Journal of Quality in Maintenance Engineering*. Vol. 12(3), pp. 239-251.

Pastor, J. T., C. A. K. Lovell. 2005. A global Malmquist productivity index. *Economics Letters*. Vol. 88(2), pp. 266-271.

Picot, A. Organisationsstrukturen im Spannungsfeld von Zentralisierung und Dezentralisierung. In: Scharfenberg H. (Ed.), *Strukturwandel in Management und Organisation: Neue Konzepte sichern die Zukunft*: Wiesbaden. 1993, pp. 217-235.

Pleshko, L. P. 2007. Strategic orientation, organisational structure, and the associated effects on performance. *Journal of Financial Services Marketing*. Vol. 12(1), pp. 53-64.

Podinovski, V. V. 2004a. Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society*. Vol. 55(12), pp. 1311-1322.

Podinovski, V. V. 2004b. Suitability and redundancy of non-homogeneous weight restrictions for measuring the relative efficiency in DEA. *European Journal of Operational Research*. Vol. 154(2), pp. 380-395.

Podinovski, V. V., A. D. Athanassopoulos. 1998. Assessing the relative efficiency of decision making units using DEA models with weight restrictions. *Journal of the Operational Research Society*. Vol. 49(5), pp. 500-508.

Portela, M. C., E. Thanassoulis. 2008. A circular Malmquist-type index for measuring productivity. *Working paper, Aston Business School research papers No. 0802.*

Portela, M. C., E. Thanassoulis, A. Horncastle, T. Maugg. 2011. Productivity change in the water industry in England and Wales: Application of the meta-Malmquist index. *Journal of the Operational Research Society*. Vol. 62(12), pp. 2173-2188.

Pourmahmoud, J. 2016. A nonlinear model for common weights set identification in network data envelopment analysis. *International Journal of Industrial Mathematics*. Vol. 8(1), pp. 87-98.

Pugh, D. S., D. J. Hickson, C. R. Hinings, C. Turner. 1968. Dimensions of organization structure. *Administrative science quarterly*. Vol. 13(1), pp. 65-105.

Qi, X.-G., B. Guo. 2014. Determining common weights in data envelopment analysis with Shannon's entropy. *Entropy*. Vol. 16(12), pp. 6394-6414.

Ramezani-Tarkhorani, S., M. Khodabakhshi, S. Mehrabian, F. Nuri-Bahmani. 2014. Ranking decision-making units using common weights in DEA. *Applied Mathematical Modelling*. Vol. 38(15-16), pp. 3890-3896.

Ramón, N., J. L. Ruiz, I. Sirvent. 2012. Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players. *Expert Systems with Applications*. Vol. 39(5), pp. 4882-4889.

Ray, S. C., E. Desli. 1997. Productivity growth, technical progress, and efficiency change in industrialized countries: Comment. *The American Economic Review*. Vol. 87(5), pp. 1033-1039.

Ray S. C., S. C. Kumbhakar, P. Dua. 2015. Benchmarking for performance evaluation: A production frontier approach. New Delhi.

Retzlaff-Roberts, D. L. 1996. Relating discriminant analysis and data envelopment analysis to one another. *Computers & Operations Research*. Vol. 23(4), pp. 311-322.

Rezaie, V., T. Ahmad, S.-R. Awang, M. Khanmohammadi, N. Maan. 2014. Ranking DMUs by calculating the interval efficiency with a common set of weights in DEA. *Journal of Applied Mathematics*. Vol. 2014.

Richmond, J. 1974. Estimating the efficiency of production. *International economic review*. Vol. 15(2), pp. 515-521.

Roll, Y., W. D. Cook, B. Golany. 1991. Controlling factor weights in data envelopment analysis. *IIE transactions*. Vol. 23(1), pp. 2-9.

Roll, Y., B. Golany. 1993. Alternate methods of treating factor weights in DEA. *Omega*. Vol. 21(1), pp. 99-109.

Roll, Y., B. Golany, D. Seroussy. 1989. Measuring the efficiency of maintenance units in the Israeli air force. *European Journal of Operational Research*. Vol. 43(2), pp. 136-142.

Ruiz, J. L., I. Sirvent. 2016. Common benchmarking and ranking of units with DEA. *Omega*. Vol. 65, pp. 1-9.

Saati, S. 2008. Determining a common set of weights in DEA by solving a linear programming. *Journal of industrial engineering international*. Vol. 4(6), pp. 51-56.

Saati, S., A. Memariani. 2005. Reducing weight flexibility in fuzzy DEA. *Applied Mathematics and Computation*. Vol. 161(2), pp. 611-622.

Sala-Garrido, R., M. Molinos-Senante, F. Hernández-Sancho. 2011. Comparing the efficiency of wastewater treatment technologies through a DEA metafrontier model. *Chemical Engineering Journal*. Vol. 173(3), pp. 766-772.

Sayed, H., R. Hamed, M. A.-G. Ramadan, S. Hosny. 2015. Using meta-goal programming for a new human development indicator with distinguishable country ranks. *Social Indicators Research*. Vol. 123(1), pp. 1-27.

Schmidt, M. 2008. The Sankey diagram in energy and material flow management – Part I: History. *Journal of Industrial Ecology*. Vol. 12(1), pp. 82-94.

Schulte-Zurhausen M. 2014. Organisation. München.

Scopus. 2018. Scopus – An eye on global research. Accessed on 21 Aug 2018. https://www.elsevier.com/__data/assets/pdf_file/0008/208772/ACAD_R_SC_FS.pdf.

Seiford, L. M., J. Zhu. 2002. Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*. Vol. 142(1), pp. 16-20.

Sherman, H. D., F. Gold. 1985. Bank branch operating efficiency: Evaluation with data envelopment analysis. *Journal of banking & Finance*. Vol. 9(2), pp. 297-315.

Sherman H. D., J. Zhu. 2006. Service productivity management: Improving service performance using data envelopment analysis (DEA). New York.

Shestalova, V. 2003. Sequential Malmquist indices of productivity growth: an application to OECD industrial activities. *Journal of Productivity Analysis*. Vol. 19(2-3), pp. 211-226.

Shirdel, G. H., S. Ramezani-Tarkhorani. 2018. A DEA-based approach for finding a favorable multi-objective shortest path. *Croatian Operational Research Review*. Vol. 9(2), pp. 149.

Si, X., L. Liang, G. Jia, L. Yang, H. Wu, Y. Li. 2013. Proportional sharing and DEA in allocating the fixed cost. *Applied Mathematics and Computation*. Vol. 219(12), pp. 6580-6590.

Sinuany-Stern, Z., L. Friedman. 1998. DEA and the discriminant analysis of ratios for ranking units. *European Journal of Operational Research*. Vol. 111(3), pp. 470-478.

SSRN. 2018. SSRN Support Page – Frequently Asked Questions. Accessed on 22 Aug 2018. https://www.ssrn.com/en/index.cfm/ssrn-faq/#what_is.

Sueyoshi, T. 1999. DEA-discriminant analysis in the view of goal programming. *European Journal of Operational Research*. Vol. 115(3), pp. 564-582.

Sueyoshi, T. 2001. Extended DEA-discriminant analysis. *European Journal of Operational Research*. Vol. 131(2), pp. 324-351.

Sun, J., J. Wu, D. Guo. 2013. Performance ranking of units considering ideal and anti-ideal DMU with common weights. *Applied Mathematical Modelling*. Vol. 37(9), pp. 6301-6310.

Sun, J., J. Wu, L. Liang, R. Y. Zhong, G. Q. Huang. 2014. Allocation of emission permits using DEA: Centralised and individual points of view. *International Journal of Production Research*. Vol. 52(2), pp. 419-435.

Sun, S. 2004. Assessing joint maintenance shops in the Taiwanese army using data envelopment analysis. *Journal of Operations Management*. Vol. 22(3), pp. 233-245.

Tangen, S. 2005. Demystifying productivity and performance. *International Journal of Productivity and Performance Management*. Vol. 54(1), pp. 34-46.

Thanassoulis, E. 1996. A data envelopment analysis approach to clustering operating units for resource allocation purposes. *Omega*. Vol. 24(4), pp. 463-476.

Thanassoulis, E. 1997. Duality in data envelopment analysis under constant returns to scale. *IMA Journal of Management Mathematics*. Vol. 8(3), pp. 253-266.

Thanassoulis, E. 1998. Estimating efficient marginal resource levels using data envelopment analysis. *The Journal of Cost Analysis*. Vol. 15(2), pp. 29-52.

Thanassoulis E. 2001. Introduction to the theory and application of data envelopment analysis: A foundation text with integrated software. New York.

Thanassoulis, E., R. G. Dyson. 1992. Estimating preferred target input-output levels using data envelopment analysis. *European Journal of Operational Research*. Vol. 56(1), pp. 80-97.

Thanassoulis, E., R. K. Shiraz, N. Maniadakis. 2015. A cost Malmquist productivity index capturing group performance. *European Journal of Operational Research*. Vol. 241(3), pp. 796-805.

Thomas, R. R., R. S. Barr, W. L. Cron, J. W. Slocum Jr. 1998. A process for evaluating retail store efficiency: A restricted DEA approach. *International Journal of Research in Marketing*. Vol. 15(5), pp. 487-503.

Tiedemann, T., T. Francksen, U. Latacz-Lohmann. 2011. Assessing the performance of German Bundesliga football players: A non-parametric metafrontier approach. *Central European Journal of Operations Research*. Vol. 19(4), pp. 571-587.

Tofallis, C. 2013. An automatic-democratic approach to weight setting for the new human development index. *Journal of Population Economics*. Vol. 26(4), pp. 1325-1345.

Toloo, M. 2012. On finding the most BCC-efficient DMU: A new integrated MIP-DEA model. *Applied Mathematical Modelling*. Vol. 36(11), pp. 5515-5520.

Toloo, M. 2013. The most efficient unit without explicit inputs: An extended MILP-DEA model. *Measurement*. Vol. 46(9), pp. 3628-3634.

Toloo, M. 2014. An epsilon-free approach for finding the most efficient unit in DEA. *Applied Mathematical Modelling*. Vol. 38(13), pp. 3182-3192.

Toloo, M., A. Kresta. 2014. Finding the best asset financing alternative: A DEA-WEO approach. *Measurement*. Vol. 55, pp. 288-294.

Toloo, M., S. Nalchigar. 2009. A new integrated DEA model for finding most BCC-efficient DMU. *Applied Mathematical Modelling*. Vol. 33(1), pp. 597-604.

Toloo, M., S. Nalchigar, B. Sohrabi. 2018. Selecting most efficient information system projects in presence of user subjective opinions: A DEA approach. *Central European Journal of Operations Research*. Vol. 26(4), pp. 1027-1051.

Toloo, M., B. Sohrabi, S. Nalchigar. 2009. A new method for ranking discovered rules from data mining by DEA. *Expert Systems with Applications*. Vol. 36(4), pp. 8503-8508.

Toloo, M., M. Tavana. 2017. A novel method for selecting a single efficient unit in data envelopment analysis without explicit inputs/outputs. *Annals of Operations Research*. Vol. 253(1), pp. 657-681.

Toloo, M., M. Tavana, F. J. Santos-Arteaga. 2017. An integrated data envelopment analysis and mixed integer non-linear programming model for linearizing the common set of weights. *Central European Journal of Operations Research*, pp. 1-18.

Tsolas, I. E. 2013. Assessing supply chain financial performance using DEA modelling. *International Journal of Business Performance and Supply Chain Modelling*. Vol. 5(3), pp. 258-271.

Tsolas, I. E. 2015. Firm credit risk evaluation: A series two-stage DEA modeling framework. *Annals of Operations Research*. Vol. 233(1), pp. 483-500.

Tsolas, I. E., V. Charles. 2015. Green exchange-traded fund performance appraisal using slacks-based DEA models. *Operational Research*. Vol. 15(1), pp. 51-77.

United Nations Development Program. 2018. The human development index - Technical notes. Accessed on 28 Aug 2018. hdr.undp.org/sites/default/files/hdr2016_technical_notes_0.pdf.

Vaninsky, A. 2006. Efficiency of electric power generation in the United States: Analysis and forecast based on data envelopment analysis. *Energy Economics*. Vol. 28(3), pp. 326-338.

Varmaz, A., A. Varwig, T. Poddig. 2013. Centralized resource planning and yardstick competition. *Omega*. Vol. 41(1), pp. 112-118.

Vassdal, T., H. M. Sørensen Holst. 2011. Technical progress and regress in Norwegian salmon farming: A Malmquist index approach. *Marine Resource Economics*. Vol. 26(4), pp. 329-341.

Vassiloglou, M., D. Giokas. 1990. A study of the relative efficiency of bank branches: An application of data envelopment analysis. *Journal of the Operational Research Society*. Vol. 41(7), pp. 591-597.

VDMA e.V. 2018. Die Deutsche Aufzugsindustrie. Accessed on 9 Nov 2018. http://auf.vdma.org/article/-/articleview/735168.

Vollmann T. E. 1988. Changing manufacturing performance measures. Boston.

Walheer, B. 2018a. Aggregation of metafrontier technology gap ratios: The case of European sectors in 1995-2015. *European Journal of Operational Research*. Vol. 269(3), pp. 1013-1026.

Walheer, B. 2018b. Cost Malmquist productivity index: An output-specific approach for group comparison. *Journal of Productivity Analysis*. Vol. 49(1), pp. 79-94.

Walheer, B. 2018c. Disaggregation of the cost Malmquist productivity index with joint and output-specific inputs. *Omega*. Vol. 75, pp. 1-12.

Wang F. 2006. Quantitative methods and applications in GIS. Boca Raton.

Wang, Q., Z. Liu, Y. Zhang. 2017. A novel weighting method for finding common weights in DEA. *Asia-Pacific Journal of Operational Research*. Vol. 34(5), pp. 1-21.

Wang, Y.-M., K.-S. Chin. 2010a. A neutral DEA model for cross-efficiency evaluation and its extension. *Expert Systems with Applications*. Vol. 37(5), pp. 3666-3675.

Wang, Y.-M., K.-S. Chin. 2010b. Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics*. Vol. 128(1), pp. 332-338.

Wang, Y.-M., K.-S. Chin, J.-B. Yang. 2007a. Measuring the performances of decision-making units using geometric average efficiency. *Journal of the Operational Research Society*. Vol. 58(7), pp. 929-937.

Wang, Y.-M., K.-S. Chin, J.-B. Yang. 2007b. Three new models for preference voting and aggregation. *Journal of the Operational Research Society*. Vol. 58(10), pp. 1389-1393.

Wang, Y.-M., P. Jiang. 2012. Alternative mixed integer linear programming models for identifying the most efficient decision making unit in data envelopment analysis. *Computers & Industrial Engineering*. Vol. 62(2), pp. 546-553.

Wang, Y.-M., Y. Luo. 2006. DEA efficiency assessment using ideal and anti-ideal decision making units. *Applied Mathematics and Computation*. Vol. 173(2), pp. 902-915.

Wang, Y.-M., Y. Luo, Y.-X. Lan. 2011. Common weights for fully ranking decision making units by regression analysis. *Expert Systems with Applications*. Vol. 38(8), pp. 9122-9128.

Webb, R. 2003. Levels of efficiency in UK retail banks: A DEA window analysis. *International Journal of the Economics of Business*. Vol. 10(3), pp. 305-322.

Weber, J., S. Gschmack. 2012. Zentralisierung von Unterstützungsprozessen: Shared Service Center für finanznahe Funktionen. *Controlling & Management*. Vol. 56(3), pp. 44-51.

Wei, Q., T.-S. Chang. 2011. Optimal profit-maximizing system design data envelopment analysis models. *Computers & Industrial Engineering*. Vol. 61(4), pp. 1275-1284.

Weinert P. 2002. Organisation: Organisationsgestaltung, Organisationsmethodik, Fallklausuren.

Wen, Y., Q. An, X. Xu, Y. Chen. 2018. Selection of Six Sigma project with interval data: Common weight DEA model. *Kybernetes*. Vol. 47(7), pp. 1307-1324.

Willem, A., M. Buelens. 2009. Knowledge sharing in inter-unit cooperative episodes: The impact of organizational structure dimensions. *International Journal of Information Management*. Vol. 29(2), pp. 151-160.

Wisner, J. D., S. E. Fawcett. 1991. Linking firm strategy to operating decisions through performance measurement. *Production and Inventory Management Journal*. Vol. 32(3), pp. 5-11.

Woodruff, R. B. 1997. Customer value: The next source for competitive advantage. *Journal of the Academy of Marketing Science*. Vol. 25(2), pp. 139.

Wu, H., S. Du, L. Liang, Y. Zhou. 2013. A DEA-based approach for fair reduction and reallocation of emission permits. *Mathematical and Computer Modelling*. Vol. 58(5-6), pp. 1095-1101.

Wu, J., J. Chu, J. Sun, Q. Zhu. 2016. DEA cross-efficiency evaluation based on Pareto improvement. *European Journal of Operational Research*. Vol. 248(2), pp. 571-579.

Wu, J., L. Liang, F. Yang, H. Yan. 2009. Bargaining game model in the evaluation of decision making units. *Expert Systems with Applications*. Vol. 36(3), pp. 4357-4362.

Yale University Center for Environmental Law & Policy. 2018. Environmental performance index: Global metrics for the environment. Accessed on 28 Aug 2018. https://epi.envirocenter.yale.edu/.

Yang, B., Y. Zhang, H. Zhang, R. Zhang, B. Xu. 2016. Factor-specific Malmquist productivity index based on common weights DEA. *Operational Research*. Vol. 16(1), pp. 51-70.

Yang, C., H.-M. Liu. 2012. Managerial efficiency in Taiwan bank branches: A network DEA. *Economic Modelling*. Vol. 29(2), pp. 450-461.

Yang, F., C. Yang, L. Liang, S. Du. 2010. New approach to determine common weights in DEA efficiency evaluation model. *Journal of Systems Engineering and Electronics*. Vol. 21(4), pp. 609-615.

Yang, F.-C., R.-H. Kao, Y.-T. Chen, Y.-F. Ho, C.-C. Cho, S.-W. Huang. 2018. A common weight approach to construct composite indicators: The evaluation of fourteen emerging markets. *Social Indicators Research*. Vol. 137(2), pp. 463-479.

Yue, P. 1992. Data envelopment analysis and commercial bank performance: A primer with applications to Missouri banks. *IC² Institute Articles.*

Zhang, M., L.-L. Wang, J.-C. Cui. 2018. Extra resource allocation: A DEA approach in the view of efficiencies. *Journal of the Operations Research Society of China*. Vol. 6(1), pp. 85-106.

Zhang, N., X. Wei. 2015. Dynamic total factor carbon emissions performance changes in the Chinese transportation industry. *Applied Energy*. Vol. 146, pp. 409-420.

Zhang, N., P. Zhou, Y. Choi. 2013. Energy efficiency, CO2 emission performance and technology gaps in fossil fuel electricity generation in Korea: A meta-frontier non-radial directional distance functionanalysis. *Energy policy*. Vol. 56, pp. 653-662.

Zhou, P., B. W. Ang, K. L. Poh. 2007. A mathematical programming approach to constructing composite indicators. *Ecological Economics*. Vol. 62(2), pp. 291-297.

Appendix

Table A1: Approaches classified to the research stream *Controlling factor weights*

Assumed management scenario	
Hybrid management	Centralized management
Afsharian et al. (2017)	Agrell and Bogetoft (2010)
Borzoei and Zohrehbandian (2013)	Ameryoun et al. (2017)
Cook et al. (2017)	Babalos et al. (2015)
Cook and Zhu (2007)	Lotfi et al. (2000)
Golany and Yu (1995)	Jahanshahloo et al. (2011)
Makui et al. (2008)	Pourmahmoud (2016)
Omrani (2013)	Qi and Guo (2014)
Roll et al. (1991)	Retzlaff-Roberts (1996)
Roll and Golany (1993)	Saati (2008)
Varmaz et al. (2013)	Saati and Memariani (2005)
Wang et al. (2007a)	Tsolas (2015)
Wu et al. (2016)	Yang and Liu (2012)
	Yang et al. (2010)

Table A2: Approaches classified to the research stream *Classification schemes*

Groups are…	*Assumed management scenario*	
	Hybrid management	Centralized management
…pre-defined	Sueyoshi (1999) Sueyoshi (2001)	
…not pre-defined	Chen (2011)	Amirteimoori and Kordrostami (2013) Hatefi and Torabi (2015)

Table A3: Approaches classified to the research stream *Resource allocation and target setting*

Assumed management scenario	
Hybrid management	Centralized management
Amirteimoori and Shafiei (2006)	Amirteimoori and Emrouznejad (2012)
Amirteimoori and Emrouznejad (2011)	Amirteimoori and Kordrostami (2005)
Aparicio and Ciurana (2006)	Amirteimoori and Tabar (2010)
Athanassopoulos (1995)	Asmild et al. (2009)
Athanassopoulos (1998)	Du et al. (2014)
Beasley (2003)	Du et al. (2010)
Bi et al. (2011)	Fang and Zhang (2008)
Färe et al. (2000)	Guedes de Avellar et al. (2007)
Hatami-Marbini et al. (2015)	Jahanshahloo et al. (2017)
Li et al. (2009)	Li et al. (2017)
Lotfi et al. (2013)	Li et al. (2018)
Nesterenko and Zelenyuk (2007)	Lozano and Villa (2004)
Thanassoulis (1996)	Lozano and Villa (2005)
Thanassoulis (1998)	Lozano et al. (2004)
Wu et al. (2009)	Lozano et al. (2011)
Zhang et al. (2018)	Mar-Molinero et al. (2014)
	Nakabayashi and Tone (2006)
	Pachkova (2009)
	Si et al. (2013)
	Sun et al. (2014)
	Wei and Chang (2011)
	Wu et al. (2013)

Table A4: Approaches classified to the research stream *Ranking of DMUs*

Assumed management scenario	
Hybrid management	Centralized management
Asosheh et al. (2010)	Amirteimoori et al. (2014)
Azizi (2011)	Azar et al. (2016)
Blancas et al. (2013)	Chen et al. (2009)
Cook et al. (1998)	Chen et al. (2018)
Doyle (1995)	Davoodi and Rezai (2012)
Foroughi (2011b)	Ebrahimnejad (2012)
Foroughi and Tamiz (2005)	Foroughi and Aouni (2012)
Gharakhani et al. (2018)	Friedman and Sinuany-Stern (1997)
Hashimoto and Wu (2004)	Jahanshahloo et al. (2010)
Kao and Hung (2005)	Jahanshahloo et al. (2005)
Liu and Peng (2008)	Kiani Mavi et al. (2010)
Ramezani-Tarkhorani et al. (2014)	Kiani Mavi et al. (2013)
Sinuany-Stern and Friedman (1998)	Liu and Peng (2009)
Toloo et al. (2009)	Makui and Momeni (2012)
Wang et al. (2017)	Ramón et al. (2012)
Wang and Chin (2010b)	Rezaie et al. (2014)
Wang and Chin (2010a)	Ruiz and Sirvent (2016)
Wang et al. (2011)	Sun et al. (2013)
Wang and Luo (2006)	Tsolas and Charles (2015)
	Wang et al. (2007b)
	Wang and Jiang (2012)

Table A5: Approaches classified to the research stream *Improving the discrimination power*

Assumed management scenario	
Hybrid management	Centralized management
Despotis (2002)	Foroughi (2012)
	Karsak and Ahiska (2005)

Table A6: Approaches classified to the research stream *Finding the (single) most efficient DMU*

Assumed management scenario	
Hybrid management	Centralized management
Toloo (2012)	Amin (2009)
Toloo and Nalchigar (2009)	Amin and Toloo (2007)
	Amin et al. (2006)
	Cook and Kress (1990)
	Foroughi (2011a)
	Shirdel and Ramezani-Tarkhorani (2018)
	Toloo (2013)
	Toloo (2014)
	Toloo and Kresta (2014)
	Toloo et al. (2017)
	Toloo et al. (2018)
	Toloo and Tavana (2017)
	Wen et al. (2018)

Table A7: Approaches classified to the research stream *Construction of composite indicators*

Assumed management scenario	
Hybrid management	Centralized management
Domínguez-Serrano and Blancas (2011)	Hatefi and Torabi (2010)
Tofallis (2013)	Sayed et al. (2015)
	Tsolas (2013)
	Yang et al. (2018)

Table A8: Approaches classified to the research stream *Dynamic performance measurement*

Assumed management scenario	
Hybrid management	Centralized management
Kao (2010)	
Yang et al. (2016)	
Afsharian and Ahn (2017)	